



NRL/MR/7320--10-9236

Validation Test Report for the Global Ocean Forecast System V3.0 – 1/12° HYCOM/NCODA: Phase II

E.J. METZGER
H.E. HURLBURT
A.J. WALLCRAFT
J.F. SHRIVER
T.L. TOWNSEND

*Ocean Dynamics and Prediction Branch
Oceanography Division*

O.M. SMEDSTAD
P.G. THOPPIL
D.S. FRANKLIN

*QinetiQ North America
Technology Solutions Group
Slidell, Louisiana*

G. PEGGION
*University of New Orleans
New Orleans, Louisiana*

February 23, 2010

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 23-02-2010		2. REPORT TYPE Memorandum Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Validation Test Report for the Global Ocean Forecast System V3.0 – 1/12° HYCOM/NCODA: Phase II				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 0603207N	
6. AUTHOR(S) E.J. Metzger, O.M. Smedstad,* P.G. Thoppil,* H.E. Hurlburt, D.S. Franklin,* G. Peggion,+ J.F. Shriver, T.L. Townsend, and A.J. Wallcraft				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 73-5094-10-5	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory Oceanography Division Stennis Space Center, MS 39529-5004				8. PERFORMING ORGANIZATION REPORT NUMBER NRL/MR/7320--10-9236	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Space & Naval Warfare Systems Command 2451 Crystal Drive Arlington, VA 22245-5200				10. SPONSOR / MONITOR'S ACRONYM(S) SPAWAR	
				11. SPONSOR / MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES *QinetiQ North America, Technology Solutions Group, Slidell, Louisiana +University of New Orleans, Work 2000 Lakeshore Drive, New Orleans, Louisiana 70148					
14. ABSTRACT Global Ocean Forecast System Version 3.0 (V3.0) is comprised of the 1/12° global HYbrid Coordinate Ocean Model (HYCOM) and the Navy Coupled Ocean Data Assimilation (NCODA) system. It is a next-generation system capable of nowcasting and forecasting the oceanic “weather,” which includes the three-dimensional ocean temperature, salinity and current structure, the surface mixed layer and the location of mesoscale features such as eddies, meandering currents and fronts. V3.0 is scheduled to replace the existing nowcast/forecast system (V2.6) based on the 1/8° Navy Coastal Ocean Model (NCOM), 1/32° Navy Layered Ocean Model (NLOM), 1/8° Modular Ocean Data Analysis System (MODAS) and NCODA. This Phase II report describes the validation testing performed on one-year hindcasts of V3.0 and V2.6. A few Phase I tasks (temperature vs. depth and acoustical proxy error analyses) have been re-evaluated along with new evaluations examining a) each system as a provider of boundary conditions to a regional nested model, b) 14-day forecast skill relative to climatology and persistence of temperature vs. depth, c) 14-day forecast skill of acoustical proxies, d) 14-day forecast skill of sea surface height and sea surface temperature, and e) a velocity comparison against glider and drifting buoy observations. Overall, this report has determined that GOFs V3.0 is performing equal to or notably better than GOFs V2.6. The superior performance of V3.0 is especially evident in providing boundary conditions to regional nested models, an important function of a global ocean nowcast/forecast system.					
15. SUBJECT TERMS HYCOM Global ocean nowcast/forecast system NCODA Model validation					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			E. Joseph Metzger
Unclassified	Unclassified	Unclassified	UL	72	19b. TELEPHONE NUMBER (include area code) (228) 688-4762

CONTENTS

1.0 INTRODUCTION	1
2.0 V3.0 SYSTEM CONFIGURATION CHANGES	2
2.0.1 Vertical remapping modifications	2
2.0.2 NCODA analysis region modifications.....	3
2.0.3 Profile assimilation time window modifications	4
2.0.4 Use of HYCOM mixed layer depth within MODAS synthetics.....	5
3.0 VALIDATION RESULTS	6
3.1 Re-evaluation of Phase I validation tasks	6
3.1.1 Temperature vs. Depth Error Analysis	6
3.1.2 Acoustical Proxy Error Analyses	9
3.2 Phase II validation tasks.....	12
3.2.1 Provision of boundary conditions to a regional nested model.....	12
3.2.2 Medium-range (14-day) forecast skill.....	17
3.2.3 Upper ocean velocity validation	25
4.0. SUMMARY AND RECOMMENDATIONS.....	28
5.0 ACKNOWLEDGEMENTS	31
6.0. REFERENCES	32
7.0 TABLE OF ACRONYMS	35

1.0 INTRODUCTION

This Validation Test Report (VTR) is the second phase of validation and verification for Global Ocean Forecast System¹ (GOFS) Version 3.0 (V3.0) that is presently running in pre-operational mode on the Cray XT5 (Einstein) at the Naval Oceanographic Office (NAVOCEANO). It is a continuation of the Phase I VTR (Metzger et al., 2008) that compares GOFS V3.0 with GOFS V2.5. The GOFS V3.0 global ocean nowcast/forecast system is based on the 1/12° global HYbrid Coordinate Ocean Model (HYCOM) and the Navy Coupled Ocean Data Assimilation (NCODA) while the GOFS V2.5 multi-model system is based on 1/8° Navy Coastal Ocean Model (NCOM), 1/32° Navy Layered Ocean Model (NLOM) and 1/8° Modular Ocean Data Assimilation (MODAS). The reader is referred to the Phase I VTR for the specifics of each system. During the time when output were being analyzed and the Phase I VTR was being written, V2.5 was the operational system at NAVOCEANO but on 24 June 2008 it was upgraded to assimilate *in-situ* profile observations via NCODA and thus became V2.6. However, a sufficiently long hindcast from V2.6 was not available at that time for validation against V3.0, and so the Phase I VTR compared V3.0 and V2.5. A year-long hindcast of V2.6 has since been integrated that spans the same analysis time frame (June 2007 - May 2008). This report re-validates a few of the tasks performed in the first VTR to examine the impact of profile assimilation in the newer GOFS V2.6, and it assesses the performance against the newer GOFS V3.0. The validation tasks include a temperature vs. depth error analysis and an acoustical proxy (mixed layer depth (MLD), sonic layer depth (SLD), below layer gradient (BLG) and deep sound channel (DSC)) error analysis. In addition, it focuses on: a) the two systems as being providers of

¹In Metzger et al. (2008) the nowcast/forecast system was referred to as the Global Ocean Prediction System (GOPS), but it has since been renamed to the Global Ocean Forecast System (GOFS). The two names can be used interchangeably between these two VTRs.

boundary conditions (BCs) to a higher resolution nested Relocatable (Relo) NCOM, b) an examination of the 14-day forecast skill relative to climatology and persistence of the temperature (T) versus depth, c) an examination of the 14-day forecast skill of the acoustical proxy variables, d) an examination of 14-day forecast skill of sea surface height (SSH) and sea surface temperature (SST), e) a comparison of the system's ability to track drifting buoys and f) the ability of the systems to simulate ocean velocity against unassimilated current measurements.

2.0 V3.0 SYSTEM CONFIGURATION CHANGES

As noted above, the Phase I VTR compares V3.0 against V2.5 (that did not assimilate *in-situ* profile observations). It shows that for the temperature vs. depth error analysis, V3.0 is performing slightly better than V2.5, but for the acoustical proxy error analysis, V2.5 generally has smaller median bias error (MdBE) and root mean square error (RMSE) than V3.0. After the V2.6 year-long hindcast was completed, we repeated these same validation tasks and discovered that *in-situ* profile assimilation improved the simulation skill of V2.6 such that V3.0 no longer produced a more accurate subsurface temperature structure. The improved simulation skill of V2.6 was mainly due to the -20 day time window of the *in-situ* profile assimilation. Because of this, both the ocean model and the assimilation methodology in V3.0 were modified and tested for improved nowcast skill and this is discussed below (sections 2.0.1 through 2.0.4). Afterward, a new V3.0 year-long hindcast was integrated and is used in this report.

2.0.1 Vertical remapping modifications

The 1/12° global HYCOM source code within GOFS V3.0 has two significant changes between the hindcasts used in the first VTR (HYCOM source code version 2.1) and this one (HYCOM source code version 2.2.18). The first is a change in the vertical remapping scheme

(remapping between z-levels and hybrid coordinates) within the hybrid grid generator (hybggen). Previously the piecewise parabolic method (PPM) was used but this is changed to a weighted essentially non-oscillatory (WENO)-like PPM scheme for increased accuracy. The second change is in hybggen when there is a too-light layer on top of a too-dense layer, i.e. when both layers would like to gain mass at the expense of each other. Previously hybggen chose each layer half of the time, but in practice the thicker of the two layers tended to gain mass and over time, the thinner layer tended to become very thin and stay that way. Now the thinner of the two layers always gains mass from the thicker layer, greatly reducing this tendency for layers to collapse. The impact of these model modifications was not examined within V3.0 due to computational expense and available manpower, but it was tested in non-assimilative mode and produced positive results. See Figure 1 for an example of the improved velocity structure in the equatorial Pacific Ocean.

2.0.2 NCODA analysis region modifications

More substantive changes are implemented in the assimilation methodology within V3.0. The first is a change in the NCODA analysis regions. Previously the globe was divided into eight regions with four elongated cyclic bands spanning approximately 22-40° of latitude in the Mercator part of the grid (67°S - 47°N) with a small overlap between adjacent regions (see Figure 13 in Metzger et al., 2008). Occasionally the NCODA analysis is inconsistent across these seams. The new scheme divides the globe into larger basin-wide regions that nominally span the Pacific, Atlantic and Indian Oceans (Figure 2) and the seams are over land masses where feasible. Avoiding narrow rectangular regions lead to a more consistent NCODA analysis of the observational data. One drawback of these new regions is the inability to fine-tune parameters used by the NCODA analysis for regions with specific dynamical characteristics, e.g. the zonally

oriented equatorial current systems (20°S-20°N). The larger regions slightly increased the computational time to compute the ocean analyses, but the small added cost is offset by increased accuracy.

2.0.3 Profile assimilation time window modifications

The following two changes to the assimilation methodology are implemented for greater consistency with V2.6 as they have been shown to improve nowcast skill in that system. One modification is to increase the time window of the *in-situ* profile observations used in assimilation from ± 12 hours to -12 days through +12 hours at the analysis time. The longer time window allows such observations to have a greater influence on successive ocean analyses, but the depth-varying weight given to each profile diminishes the older it becomes. (The time decorrelation of *in-situ* profiles is consistent between V3.0 and V2.6.) Under the old methodology, a profile is only incorporated into the analysis one time. If on the next day an altimeter pass indicated a significant SSH change, a synthetic MODAS profile would be generated that would essentially erase the previous day's observation. An observed profile should more accurately reflect the subsurface T and S structure than a synthetic profile, even if it is a few days old, thus observed profiles are now favored over synthetic profiles. The 12-day backward time window is chosen to be slightly longer than the profiling cycle of Argo floats (10 days), whereas in V2.6 the time window goes back 20 days. The following example illustrates this: with regard to Argo profiles, the new scheme in V3.0 means that ~3000 observations will be available globally on any given day to constrain the system as opposed to ~300 observations using the shorter time window. This change had the largest impact on the subsurface temperature structure in V3.0 and this will be shown below.

2.0.4 Use of HYCOM mixed layer depth within MODAS synthetics

The second modification to the assimilation methodology is the use of the ocean model's mixed layer depth (MLD) to modify and more accurately represent the vertical structure of the synthetic MODAS profiles, which tend to have a shallow MLD bias. This is similar to the methodology defined by Barron et al. (2009) which shows improved simulation skill in V2.6 when 1/32° NLOM MLD is used to modify the synthetics. In V3.0, the MODAS synthetics are made isothermal and isohaline from the surface to the depth of the forecast HYCOM MLD. Below the forecast MLD the modified synthetic relaxes back to the original T and S profile using a 10 m relaxation scale. This leads to a sharper transition at the base of the mixed layer. While V2.6 generates MODAS synthetics at every single model gridpoint (because these constitute the NCODA first guess field), V3.0 only generates a synthetic when the altimeter-derived SSH change from the previous day exceeds a user defined value (0.02 m) (because NCODA cycles with HYCOM). The modification to the MODAS synthetics had the largest impact on V3.0's ability to represent the acoustical proxy variables. Table 1 provides a synopsis of the assimilation differences between V2.5, V2.6 and V3.0 used in the Phase I and II VTRs.

Table 1: Summary of assimilation differences between V2.5, V2.6 and V3.0			
System	Time window: <i>in-situ</i> profile assimilation	MODAS synthetics	
V2.5 in Phase I VTR	No <i>in-situ</i> profile assimilation	Modified with NLOM MLD	Generated at every model gridpoint
V3.0 in Phase I VTR	-12 hrs to +12 hrs	Used as is	Only generated where significant Δ SSH
V2.6 in Phase II VTR	-20 days to +12 hrs	Modified with NLOM MLD	Generated at every model gridpoint
V3.0 in Phase II VTR	-12 days to +12 hrs	Modified with HYCOM MLD	Only generated where significant Δ SSH

3.0 VALIDATION RESULTS

3.1 Re-evaluation of Phase I validation tasks

As noted above, after the V2.6 year-long hindcast was completed, we went back and re-evaluated the temperature vs. depth and acoustical proxy error analyses. There was improved simulation skill in V2.6 compared to V2.5, and additionally it was outperforming V3.0 as set up for the Phase I validation. This was the impetus for the changes made to the HYCOM/NCODA methodology in Phase II.

3.1.1 Temperature vs. Depth Error Analysis

The temperature vs. depth error analysis is performed on both year-long V3.0 and V2.6 hindcasts (spanning June 2007 – May 2008) using profile data from the Global Ocean Data Assimilation Experiment (GODAE) server. The observations are separated into assimilated and non-assimilated profiles and only the latter type are used in the analyses presented within this report. For a given observation, the systems are sampled at the nearest model gridpoint and vertically remapped to the observation depths. For consistency with the Phase I VTR, the error analyses are broken down into the same eight regions (see Figure 13 of Metzger et al., 2008), although the actual NCODA analyses are performed over the regions defined in Figure 2. The focus areas are region MER4d that contains the major western boundary currents and the near-global region defined as MERall, which comprises the four MERxx regions. In addition some analyses focus on the western Pacific Ocean (120-170°E, 20-50°N) and Arabian Sea (45-80°E, 0-24°N) (which excludes the Red Sea and Persian Gulf). The analyses are broken into boreal seasons defined as summer (June-July-August [JJA]), fall (September-October-November [SON]), winter (December-January-February [DJF]) and spring (March-April-May [MAM]).

The same statistical metrics are used: mean error (ME), root mean square error (RMSE) and non-dimensional skill score (SS). They are defined as:

$$\begin{aligned}
 ME &= \bar{Y} - \bar{X}, \\
 RMSE &= \left[\frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2 \right]^{1/2}, \\
 R &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / (\sigma_x \sigma_y), \\
 SS &= R^2 - [R - (\sigma_y / \sigma_x)]^2 - [(\bar{Y} - \bar{X}) / \sigma_x]^2,
 \end{aligned}$$

where $\bar{X}(\bar{Y})$ and $\sigma_x(\sigma_y)$ are the mean and standard deviation of the observed (simulated) data and R is the correlation. Additional information on these statistical measures can be found in Murphy (1995). ME is the mean bias and RMSE provides the error magnitude between the model and observations. SS is a non-dimensional quantity based on correlation squared, conditional bias (middle term on the right-hand-side) and unconditional bias (right term on the right-hand-side). The conditional bias is that associated with differences in the standard deviations of the two data sets and unconditional bias is a measure of the difference in the means. A SS of 1.0 is perfect and negative skill score indicates poor performance.

The impact of the longer time window for profile assimilation and the use of model forecast MLD to modify MODAS synthetics within V3.0 is illustrated in Figures 3 and 4. A comparison of the black (V3.0 configured in the Phase I VTR) and blue (V3.0 configured with these modifications) curves shows a marked reduction in both the ME and RMSE at all depths and for all seasons. Overall, there is a significant reduction in cold bias (0.2°-0.3°C) throughout the water column and a reduction in RMSE of 0.1-0.2°C within the thermocline. The bias reduction is most evident in the upper 200-300 m for MERall (Figure 3) but extends from the surface to 500 m for region MER4d (Figure 4). The use of HYCOM MLD to modify MODAS

synthetics had the largest impact on ME in the upper 100 m of the water column, but below that the longer time window for profile assimilation reduced the bias compared to the old scheme (not shown). For most seasons the RMSE peaks between 50-200 m, which is the depth range of high variability in the thermocline with values ranging between 1.1°C and 1.5°C. At these depths, a small vertical displacement of isotherms can have large impact on the error variability. With the exception of the very near surface (top 25 m) in a few seasons, there is reduced RMSE using the newly implemented assimilation methodology. Region MER4d (Figure 4) shows seasonal variability with large RMSE (1.6°C) associated with the thermocline during JJA and SON. The erosion of the thermocline and deepening of the mixed layer (driven by winter cooling) decrease the overall error during DJF and MAM.

Figures 3 and 4 can also be used to compare V3.0 with the updated assimilation scheme (blue curves) against V2.6 that uses *in-situ* profile assimilation (red curves) and these correspond to Figures 16 and 17 from the Phase I VTR, although the profiles used in both analyses are not exactly the same. For MERall and MER4d, V3.0 generally has smaller ME from the surface down to 50-75 m. The lower temperature error at the surface in V3.0 is consistent with the findings in the Phase I VTR that showed SST bias in V2.6 to be ~0.2°C higher than V3.0 (see Tables 5-7 in Metzger et al., 2008). The smaller ME in V3.0 down to ~50-75 m may be related to a more accurate MLD used to modify the MODAS synthetics. In V2.6, the synthetics are modified using 1/32° NLOM MLD that uses a slab-type Kraus-Turner mixed layer model (Turner and Kraus, 1967), whereas in V3.0 the more sophisticated K-Profile Parameterization (KPP) mixed layer model (Large et al., 1994) is employed. Below ~100 m, the bias is small (< 0.2°C) for most seasons, although it is typically < 0.1°C for V2.6 in region MERall. With regard to the RMSE, the two systems are performing similarly. V2.6 has a small advantage between 50-

200 m in fall and winter for MERall. This is largely due to higher error at these depths in region MER3c in V3.0. Previously, when the NCODA analyses were performed over the zonally elongated subregions, the decorrelation length scales were fine-tuned in MER3c to reduce the error. Now that the analyses are performed over the basin-wide regions, decorrelation length scales appropriate for the larger basins (Pacific, Indian and Atlantic Ocean) are used and the error reduction between V3.0 in the Phase I VTR and V3.0 used here is small in MER3c. V3.0 has lower error throughout the entire water column in spring for MER4d. This is the transition season between the deep winter-time to a shallower spring-time mixed layer and thermocline, and V3.0 is responding more accurately.

3.1.2 Acoustical Proxy Error Analyses

Accurate knowledge of the underwater acoustical environment can lead to significant tactical advantages during naval operations. The 3-dimensional (3-D) T and S structure along with the upper ocean mixed layer help determine the sound speed profile that characterizes the acoustical ducts within the water column. Thus, an ocean nowcast/forecast system must be able to accurately simulate the depth of the mixed layer, i.e. the surface layer with nearly constant temperature and density vs. depth, and these other acoustical proxy variables: sonic layer depth (SLD), below layer gradient (BLG) and deep sound channel (DSC) axis. (See Figure 26 in Metzger et al. (2008) for a schematic representation of these.) MLD is defined by a 0.1°C change from the surface to a given depth. The SLD is the vertical distance from the surface to the depth of the sound speed maximum, often but not always at the base of the mixed layer. The depth of the relative minimum below the SLD is the DSC. Within the deep duct, low frequency sound energy can potentially be carried very long distances, thousands of kilometers in some cases. Lastly, BLG is defined as the sound speed rate of change with depth per 100 feet in the first 300

feet below the SLD or below the surface if the SLD is absent. As in the Phase I VTR, these quantities are derived from Naval Oceanographic Office Reference Publication 33 (RP33, 1992) with the exception that the sound speed equation by Chen and Millero (1977) and later correction by Millero and Li (1994) is used rather than that by Wilson (1960). For brevity, we show detailed analyses of MLD and SLD for regions MERall, MER4d, Arabian Sea and Western Pacific Ocean.

Because of the skewed nature of MLD and SLD, i.e. summertime values can be very shallow and wintertime values much deeper, we define these metrics

$$\text{Median Bias Error (MdB E)} = \text{median (model)} - \text{median (observation)}$$

$$\text{Median Absolute Error (MdAE)} = \text{median (|model - observation|)}$$

$$\text{Relative MdAE} = \text{MdAE (V2.6 - observation)} - \text{MdAE (V3.0 - observation)}$$

in which the prediction system quantities are validated against a common set of unassimilated observed profiles. The relative MdAE is used to determine which system is performing comparatively better, e.g. if the observed MLD = 40 m, V3.0 MLD = 42 m and V2.6 MLD = 44 m, then MdAE (V3.0 - observed) = 2 m and MdAE (V2.6 - observed) = 4 m. The relative MdAE = 2 m, so positive values indicate V3.0 has less absolute error while negative values indicate V2.6 has less absolute error.

In the Phase I VTR, V3.0 had higher MdB E for MLD and SLD and higher RMSE for all acoustic proxy variables compared to V2.5 and was the metric with the poorest performance of all those evaluated. The incorporation of HYCOM MLD to modify the MLD in the MODAS synthetics in V3.0 has improved the system's ability to more accurately represent these variables.

Using this new methodology, the shallow MLD and SLD bias has been significantly reduced in V3.0 (compare Figures 5 and 7 from this report with Figures 28 and 30 from Metzger et al., 2008). There has been little change in MLD/SLD bias between V2.5 and V2.6 due to the fact that both V2.5 and V2.6 use $1/32^\circ$ NLOM MLD-modified synthetics. Improved performance is also reflected in the relative MdAE (compare Figures 6 and 8 from this report against Figures 29 and 31 from Metzger et al., 2008). Globally, there do not appear to be any clear patterns as to one system performing regionally better than the other, but in the Arabian Sea (Western Pacific) V3.0 (V2.6) has overall smaller MdAE. For a synopsis of the overall statistics for all acoustical proxy variables, compare Tables 2-4 in this report with Tables 2-4 from Metzger et al. (2008). MdBE in V3.0 is equal to or lower than that in V2.6 for all variables with the exception of MLD for the western Pacific region. Significant improvement is especially noted for SLD and BLG indicating an improved representation of the simulated sound speed profiles. The RMSE in V2.6 is still slightly smaller than that in V3.0. This may be due to the fact that MODAS synthetics are generated at every single gridpoint in V2.6 whereas in V3.0 they are only generated when the altimeter-derived SSH change is greater than a user-defined value (2.0 cm).

Table 2: MLD/SLD/DSC/BLG Error Statistics over $\pm 50^\circ$ latitude						
	Number of unassim profiles	MdBE		RMSE		Relative MdAE (V3.0 \leq V2.6)
		V3.0	V2.6	V3.0	V2.6	
MLD	46513	-2 m	-2 m	38 m	35 m	43%
SLD	36995	-9 m	-10 m	63 m	58 m	52%
DSC	29987	-39 m	-41 m	177 m	175 m	54%
BLG	46582	-0.1 m/s/100 ft	-0.5 m/s/100 ft	1.9 m/s/100 ft	1.7 m/s/100 ft	48%

Table 3: MLD/SLD/DSC/BLG Error Statistics over the western Pacific Ocean						
	Number of unassim profiles	MdBE		RMSE		Relative MdAE (V3.0 ≤ V2.6)
		V3.0	V2.6	V3.0	V2.6	
MLD	7149	-4 m	-1 m	36 m	35 m	35%
SLD	5262	-10 m	-10 m	60 m	58 m	47%
DSC	5352	-50 m	-59 m	187 m	195 m	51%
BLG	7148	-0.1 m/s/100 ft	-0.3 m/s/100 ft	2.1 m/s/100 ft	1.8 m/s/100 ft	37%

Table 4: MLD/SLD/DSC/BLG Error Statistics over the Arabian Sea						
	Number of unassim profiles	MdBE		RMSE		Relative MdAE (V3.0 ≤ V2.6)
		V3.0	V2.6	V3.0	V2.6	
MLD	816	-2 m	-6 m	20 m	21 m	43%
SLD	552	-10 m	-10 m	25 m	21 m	53%
DSC	157	0 m	0 m	142 m	108 m	49%
BLG	815	-0.4 m/s/100 ft	-0.7 m/s/100 ft	2.2 m/s/100 ft	2.0 m/s/100 ft	50%

3.2 Phase II validation tasks

The validation tasks presented in the preceding sections are mostly re-evaluations from the Phase I VTR using the V3.0 and V2.6 systems that are shown to perform better than their previous versions. A number of additional validations tasks are performed which include (1) provision of boundary conditions to a regional nested model, (2) medium-range (14-day) forecast skill, and (3) validation of upper-ocean velocity using glider and drifter observations. They are discussed below.

3.2.1 Provision of boundary conditions to a regional nested model

One of the primary functions of GOFS V3.0 is to provide boundary conditions to higher horizontal and vertical resolution regional nested models. The grid resolution of ~9 km at the equator and ~7 km at mid-latitudes, is approximately 2-3 times coarser than that of typical Relo NCOM set-ups (Rowley et al., 2009). This 2-3x difference between the inner and outer models is

standard for many nesting configurations. In approximately the 2012 time frame, the horizontal resolution of HYCOM within GOFS will be increased to $1/25^\circ$ (~ 3.5 km at mid-latitudes) and the system will be able to directly provide BCs to nested coastal models of ~ 1 km resolution without the need for an intermediate regional nested model.

A regional Relo NCOM is configured for the area in and around Luzon Strait (118.8 - 126.3°E , 17.4 - 25.2°N) that connects the Pacific Ocean and the South China Sea and is nested within both V3.0 and V2.6. This is a severe test for the outer model to provide BCs to the inner model because of the highly energetic Kuroshio entering through the southern boundary. The inner model has 3 km resolution in both longitude and latitude and 50 vertical levels (35 σ -levels and 15 z-levels). It uses wind forcing from the 18 km resolution Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS) and thermal forcing from the 0.5° Navy Operational Global Atmospheric Prediction System (NOGAPS). It includes river discharge (Barron and Smedstad, 2002) and barotropic tidal forcing from Egbert and Erofeeva (2002) is specified at the open boundaries. The hindcasts span ~ 7 months (3 September 2007 – 31 March 2008) and observations are assimilated via NCODA with the exception of profile data. These are withheld to be used as an independent validation dataset. Outer model BCs from V3.0 and V2.6 are fed to the inner model every 6 hours along the outermost grid row/column. NCODA is turned off within a 10 point buffer zone along each sidewall.

The two Relo NCOM hindcasts have very similar solutions, but there are subtle differences noted in Figure 9 that shows the mean SSH. The strength of the Kuroshio traversing the domain from the southern to northern boundaries is stronger in the hindcast forced with V3.0 BCs, as evidenced by the tighter SSH gradient. This is also apparent by examining the meridional velocity structure near the southern boundary (Figure 10) that shows the core speed

of the Kuroshio ~ 0.1 m/s stronger and extending deeper into the water column when using V3.0 BCs. The more energetic Kuroshio within Relo NCOM results from the finer resolution of the V3.0 outer model. As noted in Metzger et al. (2008), the horizontal grid resolution is 2.2x finer in V3.0 than V2.6 and global HYCOM is able to more accurately simulate western boundary currents than global NCOM. Velocity sections east of Taiwan (Figure 11) also indicate a more energetic Kuroshio when using V3.0 BCs, although the comparison against the shipboard acoustic Doppler current profiler (ADCP) data from Liang et al. (2003) is not contemporaneous. Figure 9 also shows a more westward penetration of the Kuroshio as it enters the South China Sea (SCS) in Relo NCOM when using V2.6 BCs. Kuroshio penetration is highly non-deterministic (Metzger and Hurlburt, 2001) and the SST analysis (section 3.2.1.3) suggests a better representation in V3.0 than V2.6.

3.2.1.1 Temperature vs. depth error analysis in nested Relo NCOM

A temperature vs. depth error analysis in the upper 500 m is performed over the Luzon Strait domain using 1144 profiles (Figure 12) that were withheld from NCODA and thus are used as an independent verification dataset. Figure 13 shows a cold bias throughout the analyzed water column in the hindcast forced with V2.6 BCs, whereas in the hindcast forced with V3.0 BCs the bias is negative in the upper 125 m but then becomes positive. The bias associated with the V3.0 BC hindcast is smaller at most depths than when V2.6 outer BCs are used. A cold (warm) bias suggests weaker (stronger) advection of warm Kuroshio waters through the southern boundary. Overall, the ME in these Relo NCOM hindcasts is generally larger than that in the western Pacific region of V3.0 or V2.6 (not shown), which is approximately a constant -0.1°C to -0.2°C over the top 500 m. There is equal or smaller RMSE in the hindcast forced with V3.0 BCs over the top 450 m of the water column. The spread between these two curves is significantly

larger than noted in the analysis of V3.0 and V2.6 (Figures 3 and 4), again implying that accurate representation of the Kuroshio entering through the southern boundary is an important factor in determining simulation skill in this region.

3.2.1.2 Acoustical proxy error analysis in nested Relo NCOM

An acoustical proxy error analysis is also performed using the same set of unassimilated profiles and the results are compiled in Table 5. Spatial maps are not presented because of the non-uniform spacing of the observations across this relatively small domain. Overall, the Luzon Strait domain Relo NCOM hindcast forced with V3.0 BCs outperforms the hindcast using V2.6 BCs for all variables except the bias and MdBE of below layer gradient. This may indicate a more accurate sound speed profile in the latter. But for the other measures of the underwater environment (MLD, SLD and DSC), the hindcast using V3.0 BCs has lower bias, MdBE and RMSE.

Table 5: MLD/SLD/DSC/BLG Error Stats in Luzon Strait Relo NCOM forced with V3.0 and V2.6 Boundary Conditions							
	Mean (obs.)	ME		MdBE		RMSE	
		V3.0 BCs	V2.6 BCs	V3.0 BCs	V2.6 BCs	V3.0 BCs	V2.6 BCs
MLD	37.0 m	7.5 m	10.9 m	5.6 m	8.6 m	28.0 m	31.6 m
SLD	47.3 m	1.3 m	3.3 m	0.0 m	0.0 m	33.2 m	34.8 m
DSC	884 m	21.1 m	26.7 m	0.0 m	0.0 m	85.8 m	89.4 m
BLG	2.93 m/s/100 ft	-0.4 m/s/100 ft	0.1 m/s/100 ft	-0.4 m/s/100 ft	0.1 m/s/100 ft	1.3 m/s/100 ft	1.3 m/s/100 ft

3.2.1.3 SST error analysis in nested Relo NCOM

An SST error analysis against observed multi-channel sea surface temperature (MCSST) data is performed from the nowcast through a three-day forecast for the Luzon Strait Relo NCOM hindcasts forced with V3.0 and V2.6 BCs. (Note that through the nowcast time,

observational SST was assimilated into Relo NCOM.) The nowcast SST and n-day forecast SST valid at the nowcast time are compared against the observations. Unlike the Phase I VTR that performed a global SST analysis using numerous observation data types (drifting buoys, fixed buoys, ship observations), the relatively small size of the Relo NCOM domain and shorter time period of integration greatly restricts the actual number of observations that can be used for validation. MCSST's are the only data type with a significant number of observations (~173,000) to make for robust statistics. The next most numerous data type (drifting buoys) only had ~8800 observations and these were not evenly distributed across the Relo NCOM domain.

The mean error and RMSE from Relo NCOM relative to the MCSST data are shown in Figures 14 and 15, respectively. Overall, the Relo NCOM using V3.0 BCs has lower bias and RMSE. The bias and RMSE are relatively low in the core of the Kuroshio near the southern boundary, indicating a good representation of this current by the outer models. As noted earlier, the hindcast using V2.6 BCs has a more westward penetration of the Kuroshio into the SCS. The larger warm bias (Figure 14) in the approximate region 119-120°E, 21-23°N suggests the penetration is too deep and too much warm water is being advected into that area.

The results are also consolidated in Table 6 and the skill does not degrade from the nowcast time out through the three-day forecast. For all statistical measures, the Relo NCOM hindcast forced with V3.0 BCs has lower error/higher correlation and skill than the hindcast forced with V2.6 BCs. This is again indicative of better representation of the Kuroshio entering through the southern boundary from V3.0. The standard deviation of the observations is 1.76°C and thus the RMSE values are well below the MCSST measured variability.

Table 6: SST error statistics in Luzon Strait Relo NCOM forced by V3.0 and V2.6 BCs against ~173,000 MCSST observations								
	Mean Error (°C)		RMSE (°C)		Skill Score		Correlation	
	V3.0	V2.6	V3.0	V2.6	V3.0	V2.6	V3.0	V2.6
Nowcast	-.52	-.68	.97	1.11	.90	.89	.70	.60
1-d fcst	-.52	-.68	.97	1.11	.90	.89	.70	.60
2-d fcst	-.53	-.68	.97	1.11	.90	.89	.70	.60
3-d fcst	-.55	-.68	.97	1.11	.90	.89	.70	.60

3.2.2 Medium-range (14-day) forecast skill

A series of 14-day forecasts are used to examine medium-range forecast skill within V3.0. GOFS V2.6 was never evaluated on these times scales and thus V3.0 offers a new capability for two-week forecasts if the computational resources are available to integrate the system. Using the June 2007 – May 2008 year-long hindcast for the initial conditions, 14-day forecasts are run starting on the 1st, 8th, 15th and 22nd of each month for a total of 48 forecasts. The first five days of each ocean forecast use forecast quality NOGAPS atmospheric forcing, which is then blended toward climatological atmospheric forcing over a five day period. Beyond 10 days, purely climatological forcing is applied². No oceanic data are assimilated into the ocean model during these forecasts thus all profile data are used for validation.

3.2.2.1 Forecast skill for subsurface temperature

A temperature vs. depth error analysis is used to examine V3.0 skill as a function of forecast length and is compared against the Generalized Digital Environmental Model Version 3 (GDEM3) climatology (Carnes, 2009). The ME, RMSE and differences between V3.0 and GDEM3 are shown in Figures 16-19 for MERall, MER4d, the western Pacific and Arabian Sea regions, respectively. Metzger et al. (2008) noted that the ME and RMSE in V3.0 is not zero at

² This methodology is the same as that used by 1/32° global NLOM when it runs a 30-day forecast once per week.

the nowcast time (i.e. day 0 forecast) when compared against assimilated profiles because of errors associated with the NCODA analysis and the incremental insertion into HYCOM (their Figure 15). Thus, the error in V3.0 at the nowcast time can never be smaller than this. For the near global MERall region (Figure 16), V3.0 is clearly superior to climatology over most depths throughout most of the 14-day forecast period. Both V3.0 and GDEM3 show a cold bias over the top 500 m that does not change significantly over the forecast time frame. The near surface (top 50 m) cold bias is larger in V3.0 than GDEM3 because of excessively cold SSTs in the tropical Pacific that are impacting the near global region statistics. This will be discussed in more detail in the section on SST forecast skill. The largest bias and RMSE are seen at the base of the mixed layer and the sloping contours of V3.0 RMSE indicate the error growth rate with increasing forecast length. Regions MER4d (Figure 17) and the western Pacific (Figure 18) have similar patterns. These regions are dominated by western boundary currents with non-deterministic variability caused by mesoscale flow instabilities and since the forecasts are made without the benefit of data assimilation, the bias and RMSE are higher than for region MERall. Below ~250 m, the bias in GDEM3 is slightly less than in V3.0 over most of the forecast period but the V3.0 RMSE is smaller except near the very end. Both V3.0 and the GDEM3 climatology have highly variable bias at all depths and high RMSE in the thermocline for the Arabian Sea region (Figure 19). This region was shown to be problematic in the hindcasts analyzed in the Phase I VTR as well and is an area with poor nowcast/forecast skill. This may be due in part to a limited number of profiles available to constrain the subsurface. In addition, accurate atmospheric forcing is extremely important in this monsoon driven area. The quality of the forecast atmospheric forcing and eventual switch to climatological forcing is also contributing to the errors as a function of forecast length.

A similar temperature vs. depth analysis is performed for V3.0 and against forecasts of persisting the nowcast state over the length of the 14-day forecast. Results for the same regions are shown in Figures 20-23. An examination of the error analysis for the near global domain (MERall, Figure 20) indicates that persistence is reasonably accurate over much of forecast time frame. Part of the skill is due to the fact that this analysis region encompasses large dynamically inactive areas where the ocean environment changes very slowly. In consolidating the statistics to this level, GOFS V3.0 forecasts have a hard time beating its own persisted nowcast. However, looking at more dynamically active regions that include western boundary currents (MER4d – Figure 21, Western Pacific – Figure 22, Arabian Sea – Figure 23), the forecast skill begins to outperform persistence, especially near the end of the forecast time frame.

3.2.2.2 GOFS V3.0 forecast skill for acoustical proxy measures

The GOFS V3.0 forecast skill for MLD, SLD and BLG relative to the GDEM3 climatology and persistence of the nowcast state is depicted in Figures 24-26, respectively. Here mean error is used instead of MdBE because there are different numbers of observations used on each forecast day. An analysis is not shown for the DSC axis because there is little variation in this deep ocean measure over the 14-day forecast period. The V3.0 MLD ME is within 0-10 m over the forecast period compared to 10-20 m for the GDEM3. The V3.0 forecasts generally have positive mean error that grows slightly over the forecast period. Both MLD and SLD show a gradual decrease in the ME during the period of climatological atmospheric forcing. Depending upon the region and variable, the V3.0 forecasts have similar skill to forecasts of persistence. The Arabian Sea region tends to have the most forecast skill variability, which is due to the small number of observations used in the analysis relative to the other regions. In all cases, both V3.0 forecasts and persistence have lower bias than the GDEM3 climatology. The RMSE curves

generally increase slightly with forecast length and also exhibit a lot of variability. This is probably not because of prediction system variability but more likely due to vertical interpolation errors when the model output must be remapped to the observation depths.

3.2.2.3 SST forecast validation

In the Phase I VTR, an error analysis was performed against *in-situ* SST observations from the nowcast out through a 5-day forecast for GOFS V3.0. Upon re-examining the procedure to perform those calculations, an error was discovered. The forecast fields were not valid at the nowcast time due to a problem in the error analysis set-up. In reality, Tables 5-7 and Figure 37 in Metzger et al. (2008) show how the n-day forecasts degrade from the observations at the analysis time. Thus the SST analysis in the Phase I VTR *is correct for the nowcast time, but not for the forecast times*. This processing error has been corrected and the statistics recomputed for the hindcast and forecasts defined earlier. The results are shown in Figure 27 for all data types over the region $\pm 45^\circ$ latitude. Each of these data types has its own sources of observational error and the ship reports are the least reliable (as noted by the high RMSE values over the entire forecast period). We focus on the comparison against the MCSSTs and drifting buoys since they are the most numerous. The mean error at the nowcast time is very small (as it was in the Phase I VTR) when compared against the satellite and drifting buoy observations and it gradually becomes more negative with increasing forecast length out to about 8-10 days. It then gradually trends back toward zero until the end of the forecast period. The maximum bias growth occurs during the period when the model is forced with forecast quality NOGAPS forcing (out to 5 days). As the forecast quality forcing is blended toward climatology (days 6-10) the downward trend slows and eventually turns around during the period of purely climatological forcing. This suggests the NOGAPS forecast atmospheric forcing is the cause of the bias growth. The largest

change in the RMSE (Figure 27, right panel) is between the nowcast and the 1-day forecast and afterward there is a steady increase out to the 14-day forecast. Kara et al. (2009a) examined the atmospheric variables that drive the *seasonal* SST in non-assimilative 0.72° global HYCOM and they find that the bias and RMSE are largely controlled by the wind speed and solar radiation, although the vapor mixing ratio can also be important in the tropics (see their Figure 11 and Table 4). While it is beyond the scope of this report, additional research should be undertaken regarding the quality of the NOGAPS forecast forcing components and their impact on SST.

Figure 28 (and Tables 7 and 8) shows the same analysis but is limited to a comparison against MCSSTs and drifting buoys. It also includes the daily interpolated 4 km Pathfinder SST climatology (Kilpatrick et al., 2001) forecast error analysis and we show two other regions, the Northwest Pacific Ocean (middle column) and the Arabian Sea (right column). Over the region that covers the tropics up through the mid-latitudes, the bias in the climatology is relatively constant and slightly lower than the model forecasts after 2-3 days. However, the RMSE in the model is lower than that associated with the climatology for the entire 14-day forecast period. The Northwest Pacific region is strongly influenced by the Kuroshio and the error growth rate in V3.0 is larger than in the other regions. At the nowcast time the mean error is small in the model but rather large for the Pathfinder climatology. Comparing the error curves of the same data type, V3.0 always has smaller bias and RMSE than climatology. Lastly, in the Arabian Sea we note that the model ME rather quickly surpasses the climatological mean error, yet the RMSE in the model is generally lower or nearly equal to that associated with the Pathfinder data.

Table 7: SST error statistics vs. forecast length for GOFS V3.0 against MCSST observations over various regions. The average number of observations for the regions 45°S-45°N, NW Pacific and Arabian Sea are 32 million, 1.2 million and 1.1 million, respectively.						
Forecast length	45°S-45°N		Northwest Pacific		Arabian Sea	
	ME	RMSE	ME	RMSE	ME	RMSE
0	.02	.37	-.04	.45	-.02	.39
1	-.08	.50	-.19	.66	-.13	.50
2	-.13	.56	-.23	.72	-.18	.57
3	-.17	.61	-.33	.83	-.21	.63
4	-.19	.66	-.38	.89	-.26	.66
5	-.22	.69	-.40	.94	-.31	.70
6	-.25	.73	-.43	1.00	-.36	.76
7	-.26	.76	-.45	1.04	-.39	.80
8	-.27	.79	-.47	1.08	-.41	.84
9	-.27	.81	-.46	1.09	-.37	.82
10	-.27	.82	-.48	1.13	-.39	.86
11	-.26	.84	-.46	1.14	-.39	.82
12	-.25	.85	-.40	1.13	-.37	.81
13	-.23	.86	-.39	1.14	-.34	.79
14	-.21	.87	-.39	1.19	-.33	.82

We have also compared the GOFS V3.0 SST forecasts against persistence of the nowcast and against persistence of MODAS 2D SST analyses (Barron and Kara, 2006) (Figure 29). For the regions analyzed, the mean error is smallest at the nowcast time for the MODAS analyses and it also has the lowest bias over the entire 14-day forecast period. The bias associated with the two persistence forecasts is similar, and the V3.0 forecasts have the largest bias growth. Again, this is most likely due to errors in the NOGAPS atmospheric forecast forcing. The RMSE growth rate is similar between all forecasts but again MODAS has the lowest error at the nowcast time. In the more dynamically active Northwest Pacific region, the V3.0 forecasts have lower RMSE than the other forecasts at around one week through the end of the period, but this increases to 11 or 12 days for the other two regions. We also looked at the standard deviation of the error to

remove the bias in the error calculation (not shown). This moved the intersection of the V3.0 forecast and persisted MODAS curves 2-3 days closer to the nowcast time.

Table 8: SST error statistics vs. forecast length for GOFS V3.0 against drifting buoy observations over various regions. The average number of observations for the regions 45°S-45°N, NW Pacific and Arabian Sea are 860,000, 34,000 and 15,000, respectively.						
Forecast length	45°S-45°N		Northwest Pacific		Arabian Sea	
	ME	RMSE	ME	RMSE	ME	RMSE
0	.00	.25	.00	.35	-.05	.33
1	-.04	.45	-.07	.64	-.14	.46
2	-.06	.54	-.12	.76	-.21	.54
3	-.09	.60	-.15	.87	-.24	.58
4	-.11	.64	-.19	.91	-.29	.63
5	-.13	.68	-.24	1.01	-.33	.69
6	-.13	.72	-.25	1.04	-.36	.72
7	-.14	.75	-.24	1.08	-.37	.74
8	-.14	.77	-.23	1.14	-.36	.77
9	-.14	.80	-.23	1.17	-.36	.76
10	-.14	.81	-.19	1.22	-.35	.78
11	-.12	.83	-.17	1.26	-.36	.77
12	-.11	.83	-.17	1.26	-.36	.77
13	-.09	.84	-.15	1.24	-.38	.79
14	-.08	.86	-.12	1.25	-.36	.79

The spatial distribution of the ME and RMSE is shown in Figures 30 and 31, respectively, for forecasts of 2, 6, 10 and 14 days. The cold bias noted above spans most of the global ocean and is generally greater in the western boundary current regions. Some of this cold bias may be associated with the weaker than observed advection of warm waters by the Kuroshio and Gulf Stream. It may also be due to errors associated with the air-sea heat fluxes. In addition, note the large bias and RMSE in the eastern equatorial Pacific Ocean. One likely source of this bias is a temporal sampling bias in the shortwave radiation attenuation field (based on oceanic turbidity) used in global HYCOM. The turbidity is derived from photosynthetically active radiation derived from a climatology spanning 1997-2001 (Kara et al, 2005). Due to the large

1997-98 El Niño event, the ocean was much less turbid in this area because of reduced upwelling. A less turbid ocean means that shortwave radiation will penetrate deeper into the water column and will be spread out over a larger volume of water, leading to cooler SST. These turbidity impacts are most important near the eastern upwelling regions within a few degrees of the equator. An update of the turbidity field used in the system should lead to lower bias and RMSE in this region and potentially in other areas as well. This is currently under investigation in the 6.2 HYCOM development project.

While turbidity impacts may be contributing to the bias and RMSE near the equator, there are expansive regions with high error off the west coasts of Africa and South America away from the equator (Figures 30 and 31). Kara et al. (2009b) quantify SST error with regard to atmospheric forcing variables for the *seasonal* SST cycle and they find that air temperature, vapor mixing ratio and solar radiation are important in these regions (see their Figure 13). This suggests that additional research should be undertaken into the quality of these atmospheric forecast variables.

3.2.2.4 SSH forecast validation

Hurlburt et al. (2008) discuss model SSH forecast skill relative to the verifying analyses (their Plate 3) and we extend this to the 14-day forecasts described earlier in this section. While Hurlburt et al. (2008) also applied the analysis to “forecasts using analysis quality atmospheric forcing”, we limit the analysis to forecasts using forecast quality atmospheric forcing that blends toward climatology, since that is the way an operational system would be integrated when extended this far in time. Figure 32 illustrates median anomaly correlation and RMSE as a function of forecast length for several regions with different dynamical characteristics. In these

selected regions (and many others not shown), the V3.0 forecasts have higher anomaly correlation and lower RMSE compared to the verifying SSH analysis than do forecasts of persistence (or the RMSE of the hindcast annual mean). The Kuroshio region is chosen because it is an area of non-deterministic mesoscale flow instabilities that are insensitive to the atmospheric forcing on 14-day time scales. There the spread between V3.0 and persistence forecasts is much smaller than across the global ocean. While some of the previous analyses indicated V3.0 (and V2.6) has a somewhat difficult time depicting the subsurface structure in the Arabian Sea, this analysis indicates the sea surface is predicted with higher skill than persistence or the annual mean. Lastly, the Yellow Sea region is chosen because it is a relatively shallow sea that responds quickly to the atmospheric forcing. Forecast skill here is characterized by rapid divergence from the initial state (note the quick drop in anomaly correlation for persistence), and it is strongly determined by the timescale of accurate atmospheric forcing. The V3.0 forecast continues to have skill out to 5 days, but this too quickly drops off as the forcing blends toward climatology.

3.2.3 Upper ocean velocity validation

As of this writing, velocity observations are not assimilated into either V3.0 or V2.6, although this can be implemented within NCODA. Thus, we present an error analysis of observed versus prediction system upper ocean velocity to assess skill of the latter with respect to near surface current fields.

3.2.3.1 Validation against unassimilated gliders

During the summer/fall of 2008 and 2009, the Coastal Ocean Observing Lab at Rutgers University attempted two trans-Atlantic flights using Slocum gliders. These began off the U.S.

East coast and traveled eastward using the Gulf Stream as a tail current to speed their progress (<http://rucool.marine.rutgers.edu/atlantic/index.html>). The first flight using glider RU17 spanned 21 May – 28 October 2008 and returned average currents in the top 100 m of the ocean. The second flight using glider RU27 began on 27 April 2009 and ended 4 December 2009. It returns average currents in the top 150 m of the ocean. Because of the shallow dive extent, T and S measurements by these gliders were not assimilated into either GOFS system during the hindcasts. The gliders make several dives per day and the analysis here uses glider observations closest to 00Z which are compared to the nearest V3.0 or V2.6 model gridpoint also at 00Z and averaged over comparable depths. Figures 33 and 34 depict the glider paths and current vectors sampled once per day and the average speed for RU17 and RU27, respectively. Vector correlation $[u \cdot v / \sqrt{(u \cdot u) * (v \cdot v)}]$ and speed statistics (bias and RMS difference (RMSD)) are shown in Tables 9 and 10, respectively. Vector correlations are generally high during the time frame when the glider is within the highly energetic Gulf Stream and these fall off near the end of the period when ocean currents are weaker, although there are some exceptions. V3.0 has smaller bias than V2.6, which is too weak in comparison to both gliders. The RMSD is also smaller for V3.0 with regard to RU17, but the same for RU27. In general, these two gliders indicate GOFS V3.0 has a better representation of the upper ocean velocity field than V2.6.

Table 9: Vector correlation between unassimilated gliders and GOFS							
RU17 vs.	May	Jun	Jul	Aug	Sep	Oct	All
V3.0	.48	.65	.64	.58	.26	.01	.58
V2.6	.33	.28	.63	.34	-.24	.41	.32
RU27 vs.							
V3.0	.48	.48	.61	.19	.56	---	.47
V2.6	.58	.53	.33	.63	.60	---	.49

Table 10: Speed statistics between unassimilated gliders and GOFS		
RU17 vs.	Bias (cm/s)	RMSD (cm/s)
V3.0	2.7	28.8
V2.6	-6.5	34.7
RU27 vs.		
V3.0	-4.2	29.8
V2.6	-12.1	29.8

3.2.3.2 Validation against unassimilated drifting buoys

Lastly, we compare the V3.0 and V2.6 upper ocean velocity field against a set of drifting buoys obtained from the National Oceanic and Atmospheric Administration (NOAA) Global Drifter Program (GDP) similar to the analysis of Barron et al. (2007). Observed SST from these drifters is assimilated into V3.0, but the velocities are not and thus they are an independent validation data set for this analysis. The drifters are drogued at 15 m and both instantaneous V2.6 and V3.0 velocity output are interpolated to this depth. Once per day, drifter trajectories are computed using the interpolated V2.6 and V3.0 velocity snapshots. The initial position of each of these trajectories corresponds with an observer drifter, and the velocity snapshots are used to advect the drifters for the next seven days. Median separation error is used as a measure of the degree to which the observed and predicted trajectories diverge as a function of time. Initially this analysis was performed globally using output from the two systems. However, the temporal sampling frequency of the prediction system velocity fields has a large impact on the separation statistics (C. Barron, personal communication). Using the relatively coarse 1-day velocity fields from V2.6 and V3.0 to compute the trajectories, the separation distance as a function of time is generally larger than the results from Barron et al. (2007), although it must be noted that the analysis regions are not exactly the same and different time frames are used. In hindsight, higher temporal frequency output should have been saved globally to properly make this comparison, but it was not done because of the large storage requirements.

Six-hourly output from GOFS V3.0 is available over the Luzon Strait region since it provided boundary conditions for the Relo NCOM analysis in section 3.2.1. While this is a small region and will not provide a global analysis, it can be used to examine the impact of the temporal sampling. Figure 35 shows median separation distance as a function of time for those drifters within the Luzon Strait domain. The solid (dashed) curves use 6-hourly (daily) velocity fields to compute the trajectories and additionally persistence of the initial observed drifter position (i.e. no drift) is shown. Only those drifting buoys that remained totally within the domain are used in this analysis. Using the higher frequency velocity output (solid lines) the two prediction systems perform similarly but V3.0 has a slight edge between two and four days. However, the small sample size (97 trajectory pairs) prevents us from saying this is a robust result. Nonetheless, prediction system results are clearly superior to using persistence. Comparing the solid and dashed curves illuminates an interesting result. There is essentially no impact of using higher temporal sampling for V2.6 but a positive impact for V3.0. The SSH variability in this region (not shown) is substantially higher in V3.0 than in V2.6 indicating more mesoscale variability in the former system. This implies more eddies and variations associated with the Kuroshio as it traverses the analysis domain. Using higher temporal frequency sampling in V3.0 thus leads to smaller separation distances as more accurate velocities are used to compute the trajectories.

4.0 SUMMARY AND RECOMMENDATIONS

Between the Phase I VTR and this Phase II VTR, there has been continued effort to improve the assimilation methodology in GOFS V3.0 via the 6.4 Ocean Data Assimilation and Large Scale Prediction projects. The two main changes are 1) increasing the time window of *in-situ* profile assimilation to -12 days through + 12 hours and 2) implementing HYCOM MLD to

modify the MODAS synthetics used to project the surface information down into the water column. Both changes have led to improved prediction system skill with regard to error analyses of temperature vs. depth and the underwater acoustical environment (which have been re-evaluated and compared to output from the Phase I VTR). Although V2.6 used a longer time window of profile assimilation (-20 days through + 12 hours) and MODAS MLD modification from a different source ($1/32^\circ$ global NLOM), the subsurface temperature structure in both V2.6 and V3.0 is very similar. It should be noted that on a given day the number of profile observations and MODAS synthetics that go into V2.6 is significantly higher than what goes into V3.0. Thus, GOFS V3.0 is able to dynamically evolve more freely and is less constrained by the climatologically-derived MODAS synthetics. Typically V3.0 has smaller bias in the top 50 m for all the regions examined and this improvement is likely associated with the second assimilation modification mentioned above. Otherwise, the two systems are performing nearly equally although each can be found to have lower bias and RMSE in certain regions and seasons. In all cases, the recently modified V3.0 and V2.6 outperform the original V3.0 and V2.5 systems. The largest improvement in this version of GOFS V3.0 has been with regard to the underwater acoustical environment. Whereas the Phase I VTR indicated V2.5 was outperforming V3.0, a re-analysis in this report indicates the improved V3.0 is equal to or better than V2.6 with regard to MdB. The largest improvements came with regard to the sonic layer depth and below layer gradient.

A primary function of global ocean prediction systems is to provide BCs to regional nested models and this has been tested using 3 km Relocatable NCOM for the area surrounding Luzon Strait in the western Pacific. This is a severe test for an outer model to provide boundary conditions because of the highly energetic Kuroshio entering through the southern boundary. The

Relo NCOM hindcast using V3.0 BCs has a better representation of the Kuroshio at the southern boundary and along the east coast of Taiwan than the hindcast using V2.6 BCs. In addition, the temperature vs. depth, acoustical proxy and SST error analyses are generally all better in the hindcast using V3.0 BCs. Thus, V3.0 is superior to V2.6 in providing boundary conditions to Relo NCOM in the region tested.

The medium-range forecast skill of V3.0 using 48 14-day forecasts is also examined. The subsurface temperature and acoustical proxy measures are examined relative to the GDEM3 climatology and to forecasts of persistence. GOFS V3.0 is generally better than climatology over most depths and regions. However, forecasts of persistence are often essentially the same as model forecasts. This occurs because the large analysis regions include large areas where the changes are of low amplitude. In more dynamically active regions that include the western boundary currents, forecast skill clearly outperforms persistence, especially near the end of the forecast time frame. An SST error analysis indicates the bias grows to values of only -0.27°C or -0.14°C at $\sim 8\text{-}10$ days when compared against MCSST and drifting buoys, respectively. However, after the model forcing returns to climatology, the bias trends back toward zero. This suggests the NOGAPS atmospheric forecast forcing is the cause of the bias drift. The blending of forecast quality NOGAPS forcing with climatology can have a significant impact on the forecasts error, especially during the period of propagating storms and warm/cold atmospheric fronts. Lastly, an SSH anomaly correlation analysis indicates that V3.0 has skill over forecasts of persistence in all the regions examined.

Upper ocean velocity is examined by comparing the two prediction systems against unassimilated gliders and drifting buoys. With regard to the gliders, V3.0 had lower speed biases and RMSD than V2.6, although this is based on only two samples. A global analysis using

drifting buoys highlighted the requirement to use prediction system output with a temporal frequency of greater than once per day, especially in regions with meandering fronts and numerous mesoscale eddies.

Overall, this report has determined that GOFS V3.0 is performing equal to or notably better than GOFS V2.6. The superior performance of V3.0 is especially evident in providing boundary conditions to regional nested models, an important function of a global ocean nowcast/forecast system.

5.0 ACKNOWLEDGEMENTS

This work was funded as part of the NRL 6.4 Large Scale Prediction and 6.4 Ocean Data Assimilation projects, managed by the Space and Naval Warfare Systems Command under program element 0603207N. The numerical simulations were performed on the NAVOCEANO IBM-Power 5+ and Power 6 and the Cray XT5 at Stennis Space Center, Mississippi using grants of computer time from the Department of Defense High Performance Computing Modernization Program. The Validation Test Panel consisted of: Frank Bub, Chris DeHaan, Andrea Mask and Bruce Lunde (NAVOCEANO), John Harding (Northern Gulf Institute), Joe Metzger and Harley Hurlburt (NRL). The authors would like to thank the various NRL contributors for their support in global HYCOM/NCODA development and validation. Among them are James Cummings for continued development of NCODA and Bob Helber for the software and expert advice with the MLD/SLD/DSC/BLG calculations. We also thank John Kerfoot at Rutgers University for supplying the RU17 and RU27 glider data.

6.0 REFERENCES

- Barron, C.N., R.W. Helber, L.F. Smedstad, T.L. Townsend and J.M. Dastugue, 2009: Validation Test Report for GOFS 2.6: MLD-modified synthetics and NCODA profile assimilation in global NCOM. *NRL Memo. Report*. (in preparation).
- Barron, C.N. and A.B. Kara, 2006: Satellite-based daily SSTs over the global ocean. *Geophys. Res. Lett.*, **33**, L15603, doi:10.1029/2006GL026356.
- Barron, C. N. and L.F. Smedstad, 2002: Global river inflow within the Navy Coastal Ocean Model. *Proc. Oceans 2002 MTS/IEEE Conference*, Biloxi, Mississippi, USA, pp. 1472–1479.
- Barron, C. N., L.F. Smedstad, J.M. Dastugue and O.M. Smedstad, 2007: Evaluation of ocean models using observed and simulated drifter trajectories: Impact of sea surface height on synthetic profiles for data assimilation. *J. Geophys. Res.*, **112**, C07019, doi:10.1029/2006JC003982.
- Carnes, M.R., 2009: Description and evaluation of GDEM-V3.0. NRL Memorandum Rpt., NRL/MR/7330—09-9165.
- Chen, C.T. and F.J. Millero, 1977: Speed of sound in seawater at high pressures. *J. Acoust. Soc. Am.*, **62**, 1129-1135.
- Egbert, G. and S. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Ocean Technol.*, **19**, 183-204.
- Hurlburt, H.E., E.P. Chassignet, J.A. Cummings, A.B Kara, E.J. Metzger, J.F. Shriver, O.M. Smedstad, A.J. Wallcraft and C.N. Barron, 2008: Eddy-resolving global ocean prediction. In

- Ocean Modeling in an Eddying Regime, Geophysical Monograph 177, M. Hecht and H. Hasumi, eds., American Geophysical Union, Washington, DC, pp. 353-381.
- Johnson, G.C., B.M. Sloyan, W.S. Kessler, and K.E. McTaggart, 2002: Direct measurements of upper ocean currents and water properties across the tropical Pacific Ocean during the 1990's. *Prog. Oceanogr.*, **52**, 31-61.
- Kara, A.B., A.J. Wallcraft and H.E. Hurlburt, 2005: A new solar radiation penetration scheme for use in ocean mixed layer studies: An application to the Black Sea using a fine-resolution Hybrid Coordinate Ocean Model (HYCOM). *J. Phys. Oceanogr.*, **35**, 13-32.
- Kara, A.B., A.J. Wallcraft, H.E. Hurlburt and W.-Y. Loh, 2009a: Which surface atmospheric variable drives the seasonal cycle of sea surface temperature over the global ocean? *J. Geophys. Res.*, **114**, D05101, doi:10.1029/2008JD010420.
- Kara, A.B., A.J. Wallcraft, H.E. Hurlburt and W.-Y. Loh, 2009b: Quantifying SST errors from an OGCM in relation to atmospheric forcing variables. *Ocean Modelling*, **29**, doi:10.1016/j.ocemod.2009.03.001.
- Kilpatrick, K. A., G. P. Podesta, and R. Evans, 2001: Overview of the NOAA/NASA Advanced Very High Resolution Radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, 9179– 9197.
- Large, W.G., J.C. Mc Williams and S.C. Doney, 1994: Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* **32**, 363-403.
- Liang, W.D., T.Y. Yang, M.T. Ko and W.S. Chuang, 2003: Upper-ocean currents around Taiwan, *Deep Sea Res. II*, **50(6-7)**, 1085-1105.

- Metzger, E.J. and H.E. Hurlburt, 2001: The nondeterministic nature of Kuroshio penetration and eddy shedding in the South China Sea. *J. Geophys. Res.*, **31**, 1712-1732.
- Metzger, E.J., O.M. Smedstad, P. Thoppil, H.E. Hurlburt, A.J. Wallcraft, D.S. Franklin, J.F. Shriver and L.F. Smedstad, 2008: Validation Test Report for the Global Ocean Prediction System V3.0 - 1/12° HYCOM/NCODA: Phase I. *NRL Memo. Report*, NRL/MR/7320--08-9148.
- Millero, F.J. and X. Li, 1994: Comments on “On equations for the speed of sound in seawater”. *J. Acoust. Soc. Am.*, **95**(5), 2757-2759.
- Murphy, A.H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting*, **10**, 681-688.
- Rowley, C., P.J. Martin and J.A. Cummings, 2009: The Naval Research Laboratory Relocatable Ocean Nowcast/Forecast System. *U.S. Navy Journal of Underwater Acoustics*. (submitted).
- RP33: Fleet Oceanographic and Acoustic Reference Manual, 1992: Naval Oceanographic Office Reference Publication 33, Naval Oceanographic Office, Stennis Space Center, MS.
- Turner, J.S. and E.B. Kraus, 1967: A one-dimensional model of the seasonal thermocline II: The general theory and its consequences. *Tellus*, **19**, 98-105.
- Wilson, W.D., 1960: Equation for the speed of sound in sea water. *J. Acoust. Soc. Amer.*, **32**, 1357.

7.0 TABLE OF ACRONYMS

ADCP	Acoustic Doppler Current Profiler
BCs	Boundary Conditions
BLG	Below Layer Gradient
COAMPS	Coupled Ocean Atmosphere Mesoscale Prediction System
DSC	Deep Sound Channel
GDEM3	Generalized Digital Environmental Model Version 3
GDP	Global Drifter Program
GODAE	Global Ocean Data Assimilation Experiment
GOFS	Global Ocean Forecast System
GOPS	Global Ocean Prediction System
HYCOM	HYbrid Coordinate Ocean Model
KPP	K-Profile Parameterization
MCSST	Multi-Channel Sea Surface Temperature
MdAE	Median Absolute Error
MdBE	Median Bias Error
ME	Mean Error
MLD	Mixed Layer Depth
MODAS	Modular Ocean Data Analysis System
NAVOCEANO	Naval Oceanographic Office
NCODA	Navy Coupled Ocean Data Assimilation
NCOM	Navy Coastal Ocean Model
NLOM	NRL Layered Ocean Model
NOAA	National Oceanic and Atmospheric Administration
NOGAPS	Navy Operational Global Atmospheric Prediction System
NRL	Naval Research Laboratory
PPM	Piecewise Parabolic Method
RMSE	Root Mean Square Error
S	Salinity
SCS	South China Sea
SLD	Sonic Layer Depth
SS	Skill Score
SSH	Sea Surface Height
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
T	Temperature
VTR	Validation Test Report
WENO	Weighted Essentially Non-Oscillatory

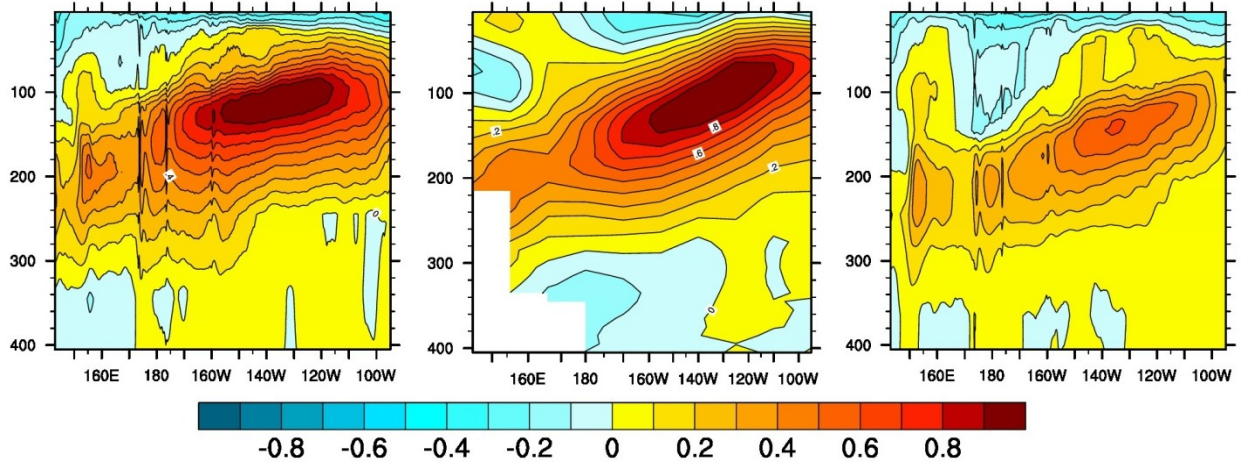


Figure 1: Zonal velocity (m/s) along the equator in the Pacific Ocean in the top 405 m from non-assimilative 1/12° global HYCOM using the new hybgen (left), observations at the Tropical Atmosphere Ocean array (Johnson et al., 2002) (middle), and non-assimilative 1/12° global HYCOM using the old hybgen (right).

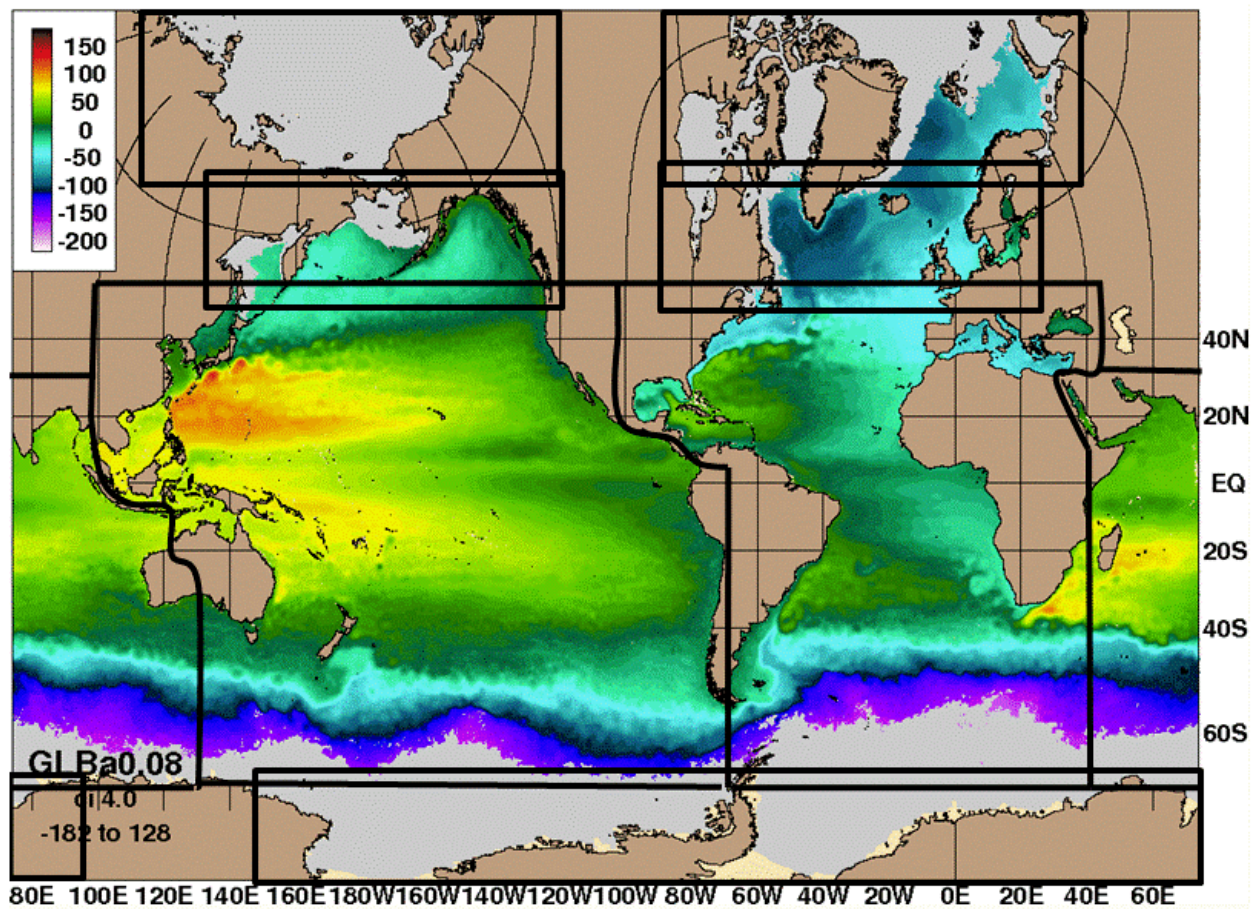


Figure 2: The NCODA analysis regions (defined by the black lines) used in GOFS V3.0 overlain on a field of sea surface height. In the Mercator part of the grid, these nominally span the Pacific, Atlantic and Indian Oceans.

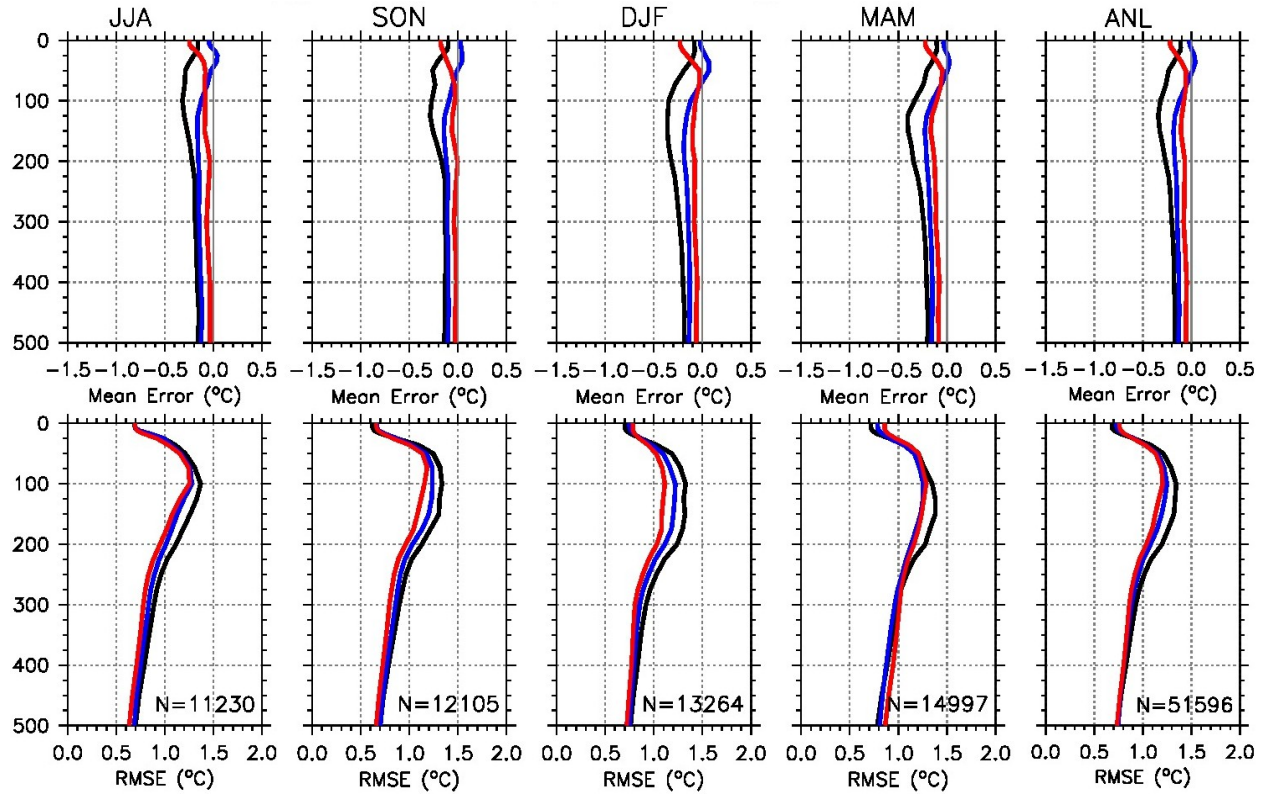


Figure 3: Temperature (°C) vs. depth error analysis in the upper 500 m against unassimilated profiles for the near-global MERall region for the four seasons – first column = summer (JJA), second column = fall (SON) ... last column = annual (ANL). The top row is mean error and the bottom row is RMSE. The black curves represent V3.0 as configured in the Phase I VTR, the blue curves represent V3.0 as configured in this report and the red curves represent V2.6. The number of unassimilated profiles used in each season is indicated by N = xxxxx in the bottom row and each simulation is analyzed against a common set of observations. This is the number of near-surface observations used and the value decreases with depth since not all profiles extend down to 500 m depth.

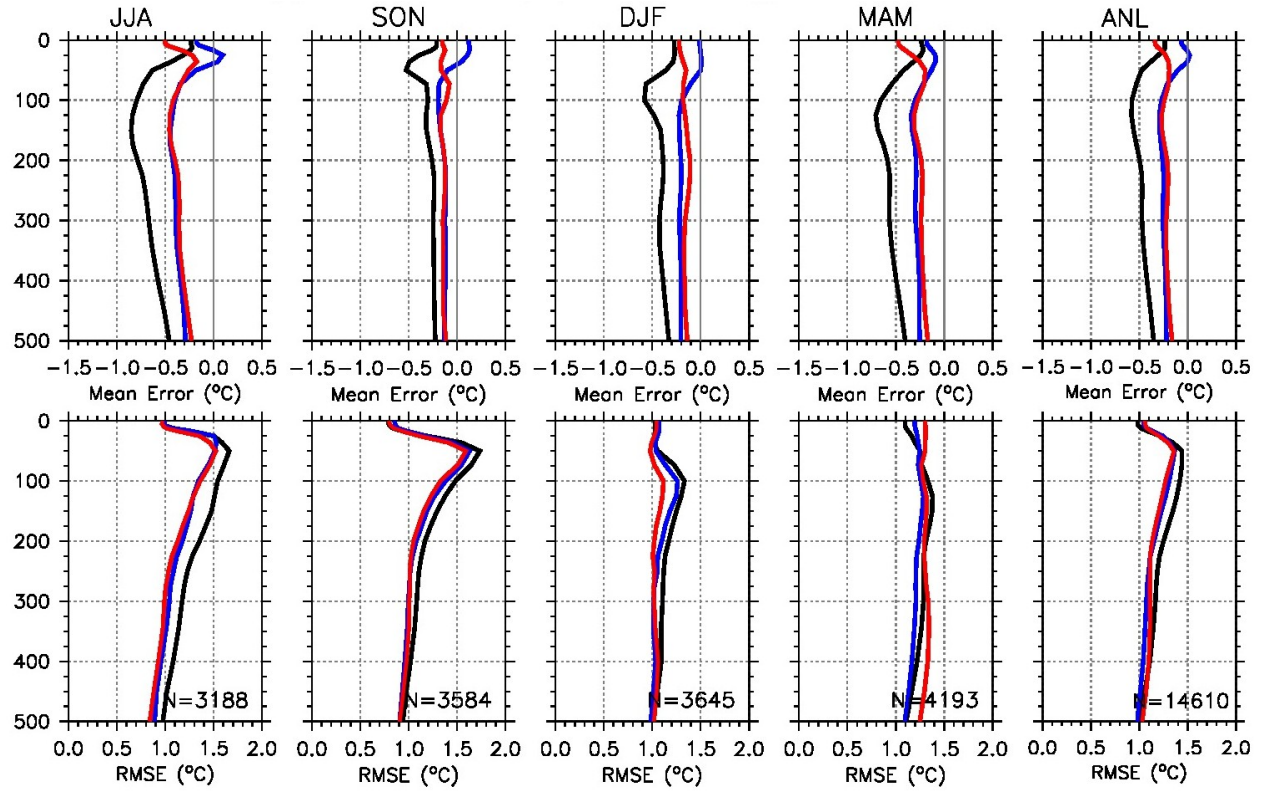


Figure 4: As in Figure 3 except for region MER4d (that includes the western boundary currents).

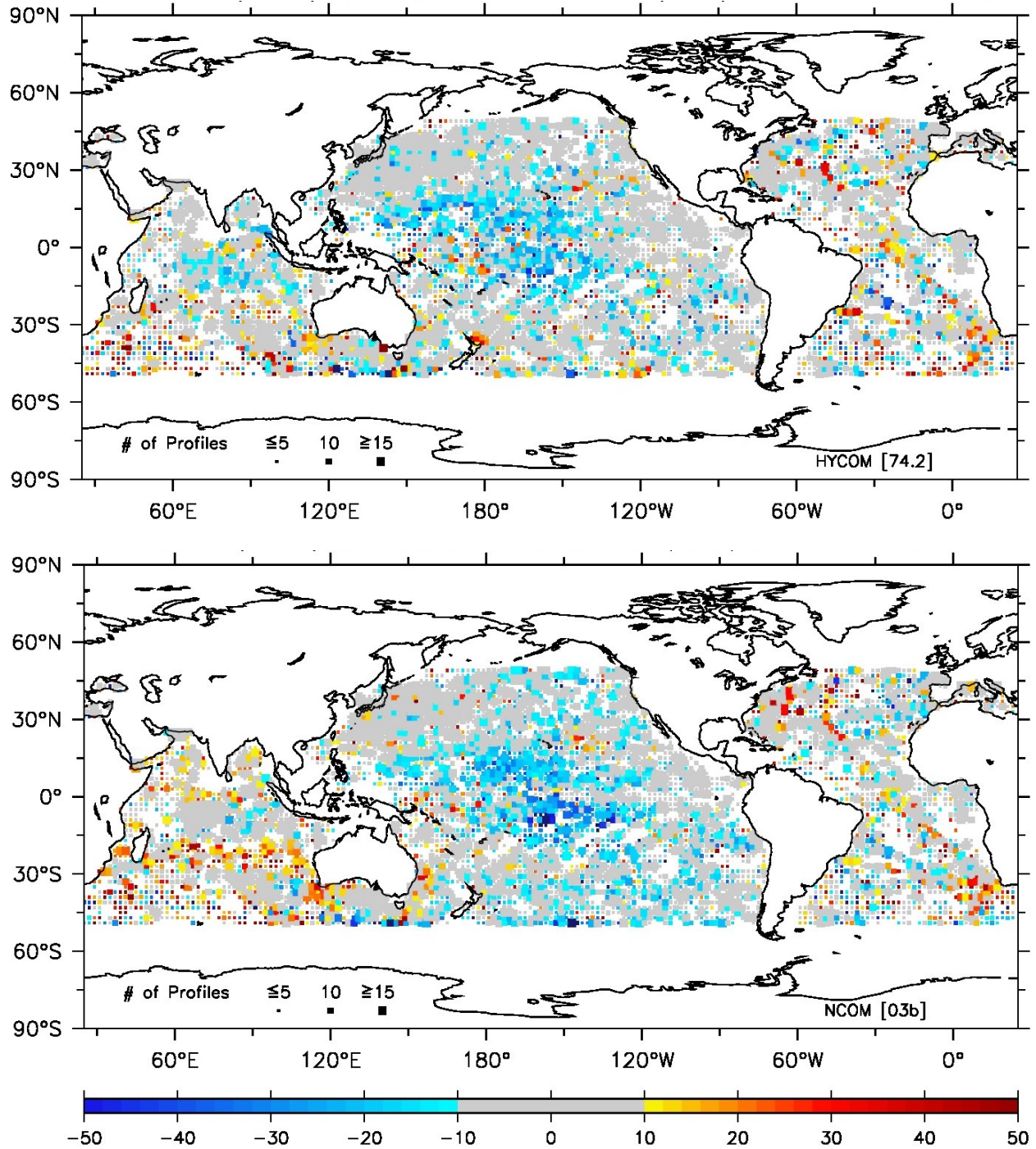


Figure 5: MLD Median Bias Error (MdBE) for V3.0 (top) and V2.6 (bottom) against 46513 unassimilated profiles over the hindcasts spanning 1 June 2007-31 May 2008. For V3.0, the basin average MdBE = -2 m, RMSE = 38 m and 60% of the points are within 10 m of the observation. For V2.6, the basin average MdBE = -2 m, RMSE = 35 m and 58% of the points are within 10 m of the observation. Red values indicate the nowcast MLD is deeper than observed while blue values indicate it is shallower. The analysis is limited to $\pm 50^\circ$ latitude. The data are averaged over 2° bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend within Antarctica.

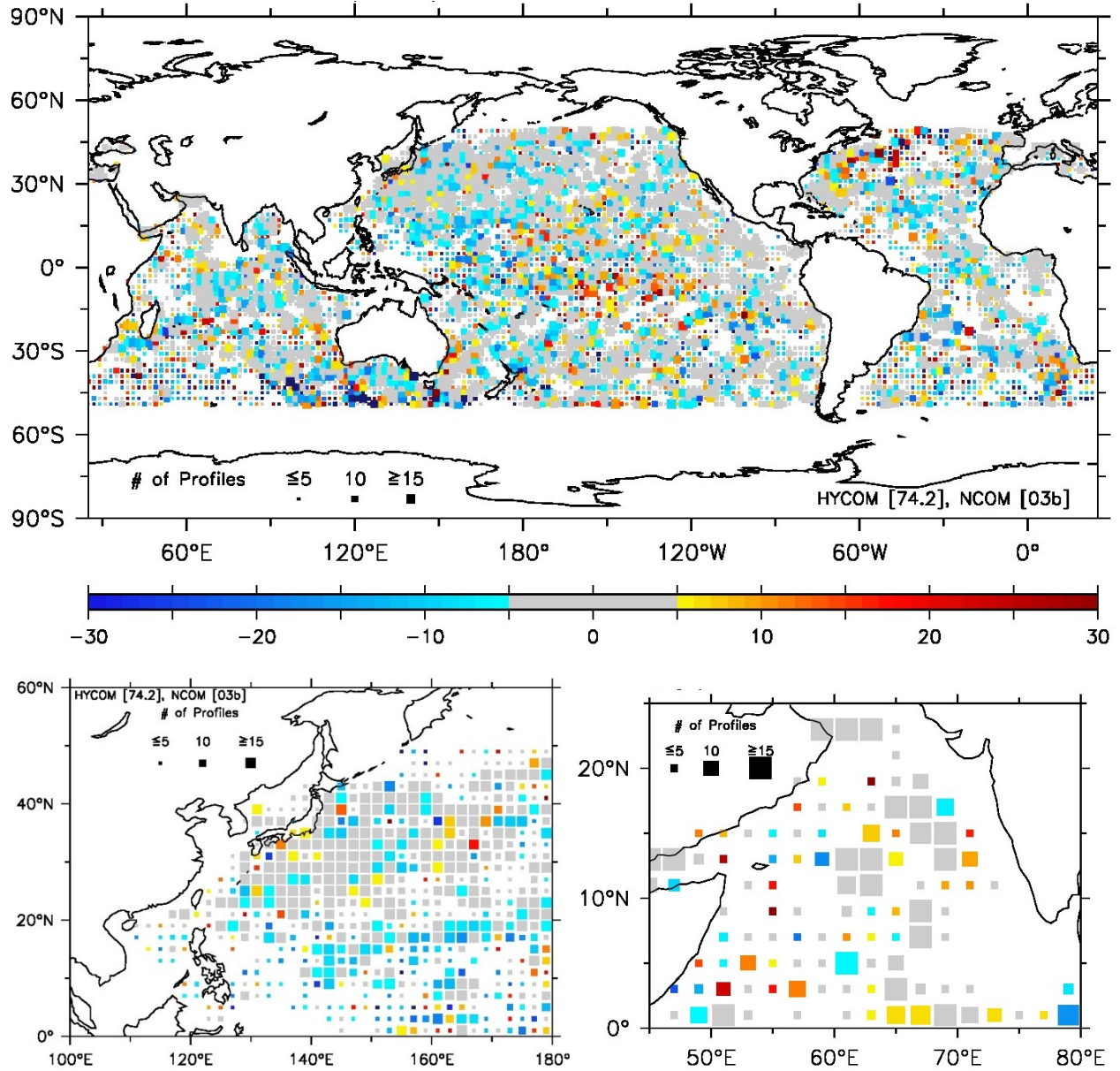


Figure 6: MLD relative Median Absolute Error (MdAE) for the whole domain (top) using 46513 unassimilated profiles, western Pacific Ocean (bottom left) using 7149 unassimilated profiles and the Arabian Sea (bottom right) using 816 unassimilated profiles. Positive (negative) values indicate V3.0 (V2.6) has lower absolute error. The analysis in the top panel is limited to $\pm 50^\circ$ latitude. MdAE is less than 5 m in those boxes colored gray. The data are averaged over 2° bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend. The percentage of points where V3.0 has equal or less absolute error than V2.6 is 43% for the whole domain, 35% in the western Pacific and 76% in the Arabian Sea.

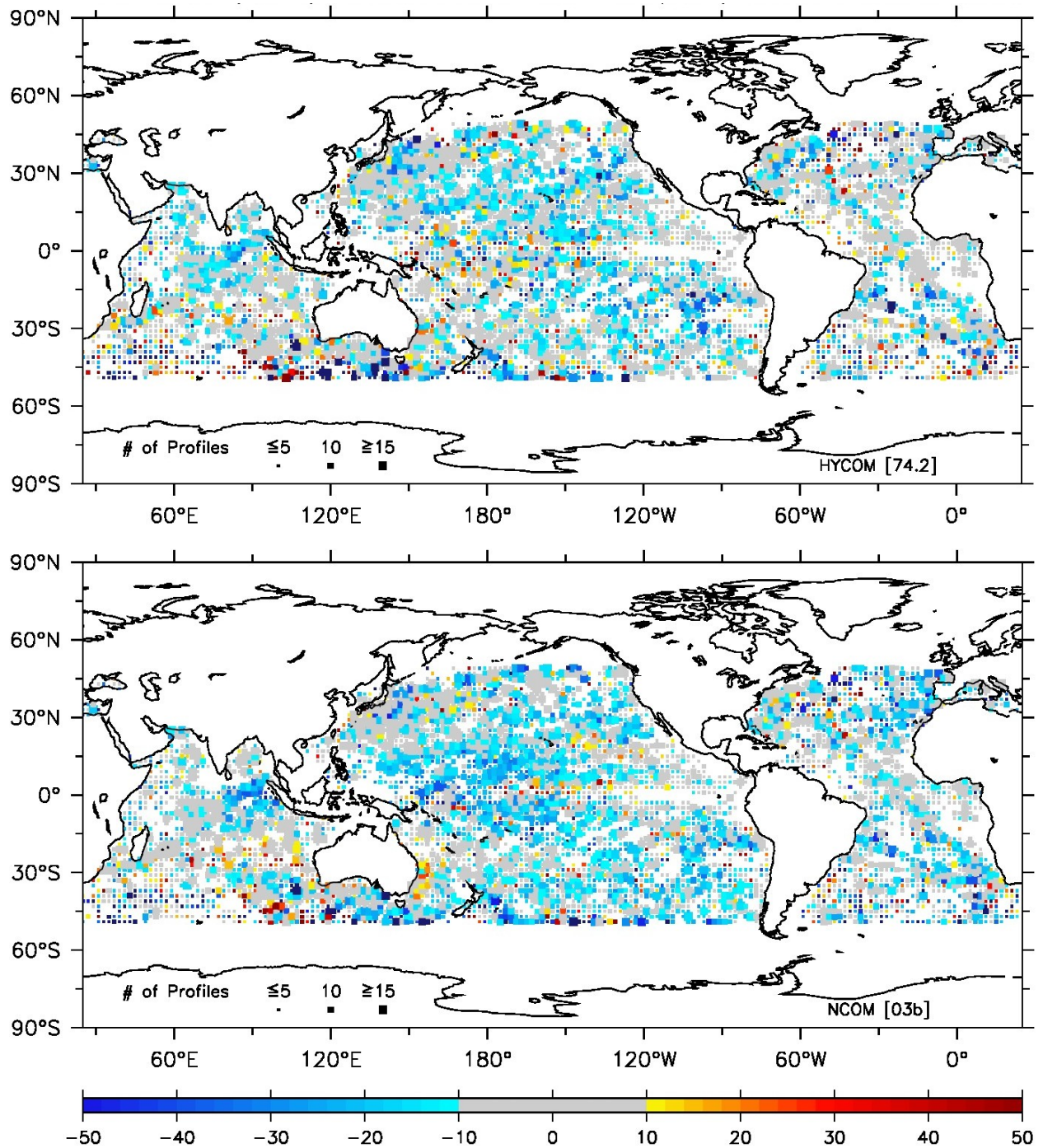


Figure 7: SLD Median Bias Error (MdBE) for V3.0 (top) and V2.6 (bottom) against 36995 unassimilated profiles over the hindcasts spanning 1 June 2007-31 May 2008. For V3.0, the basin average MdBE = -9 m, RMSE = 63 m and 41% of the points are within 10 m of the observation. For V2.6, the basin average MdBE = -10 m, RMSE = 58 m and 37% of the points are within 10 m of the observation. Red values indicate the nowcast MLD is deeper than observed while blue values indicate it is shallower. The analysis is limited to $\pm 50^\circ$ latitude. The data are averaged over 2° bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend within Antarctica.

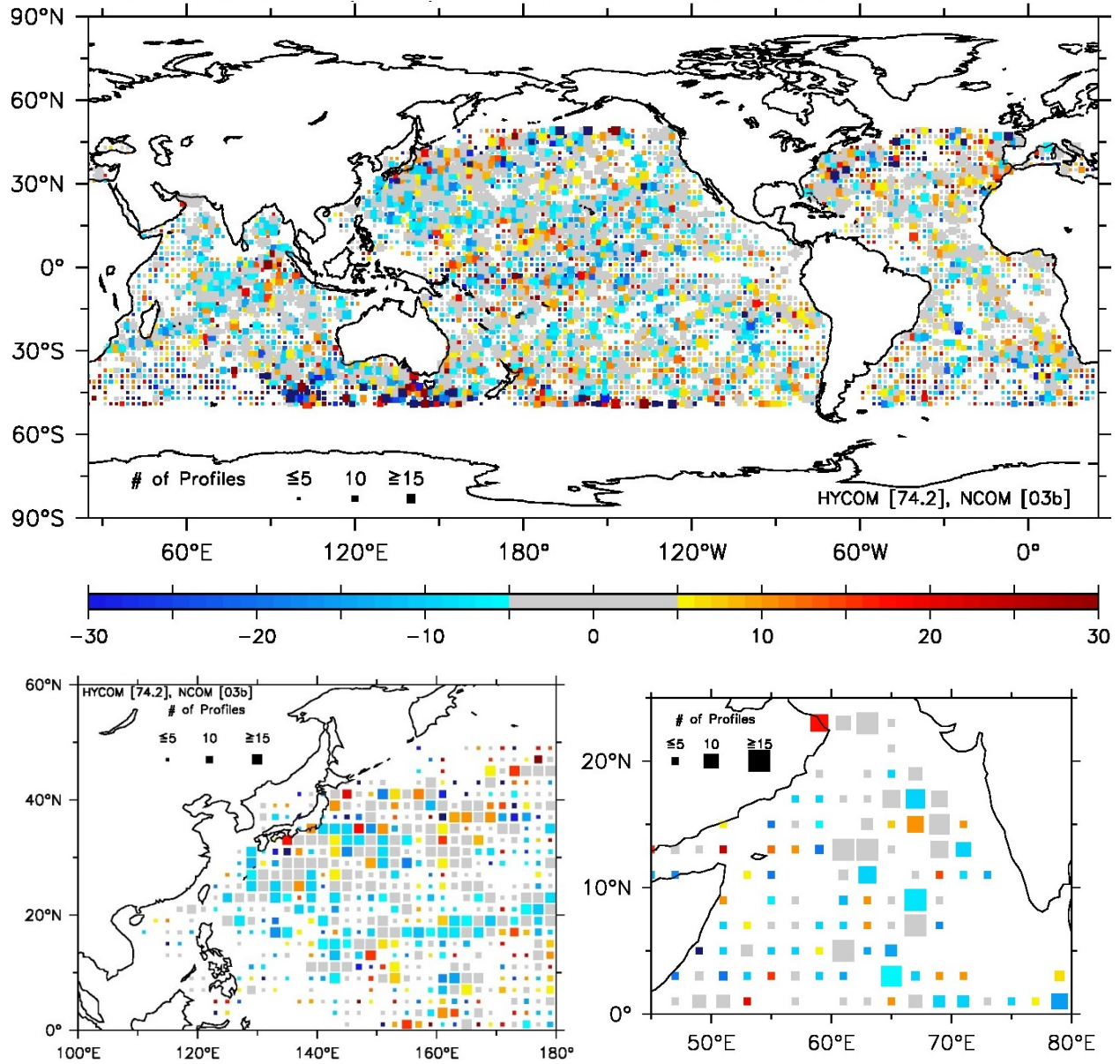


Figure 8: SLD relative Median Absolute Error (Mdae) for the whole domain (top) using 36995 unassimilated profiles, western Pacific Ocean (bottom left) using 5262 unassimilated profiles and the Arabian Sea (bottom right) using 552 unassimilated profiles. Positive (negative) values indicate V3.0 (V2.6) has lower absolute error. The analysis in the top panel is limited to $\pm 50^\circ$ latitude. Mdae is less than 5 m in those boxes colored gray. The data are averaged over 2° bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend. The percentage of points where V3.0 has equal or less absolute error than V2.6 is 52% for the whole domain, 47% in the western Pacific and 53% in the Arabian Sea.

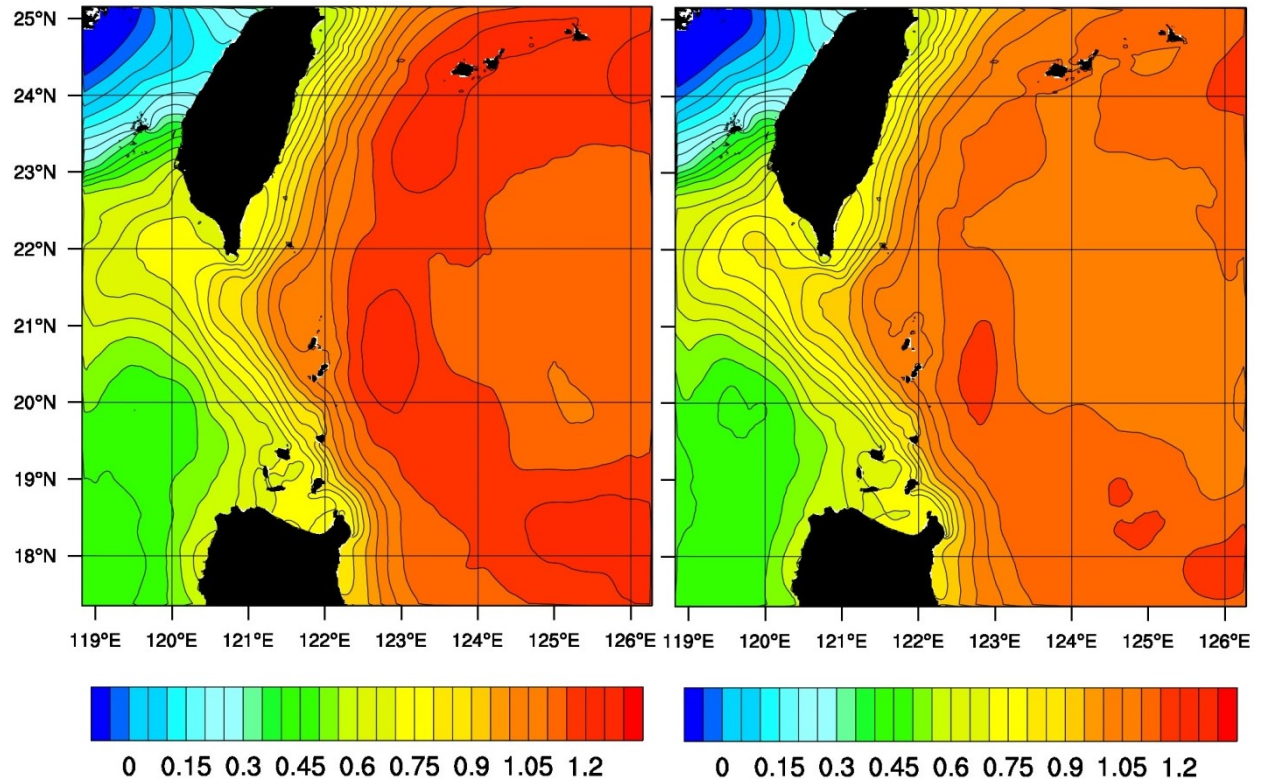


Figure 9: Mean sea surface height (in meters) from the Relo NCOM system configured for the area in and around Luzon Strait forced with BCs from V3.0 (left) and V2.6 (right). The mean is computed over the time frame 3 September 2007 – 31 March 2008. The contour interval is .05 m.

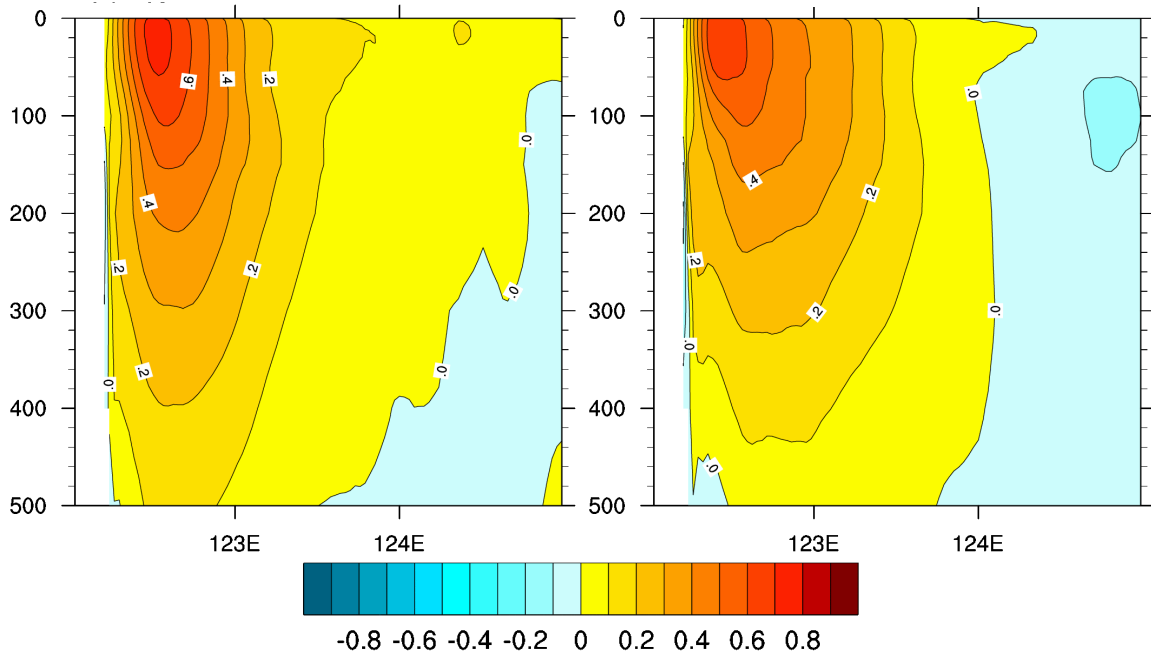


Figure 10: Mean meridional velocity (m/s) from the surface to 500 m at 18°N from the Taiwan coast to 125°E in the Relo NCOM system configured for the area in and around Luzon Strait forced with BCs from V3.0 (left) and V2.6 (right). The mean is computed over the time frame 3 September 2007 – 31 March 2008. The contour interval is .1 m/s.

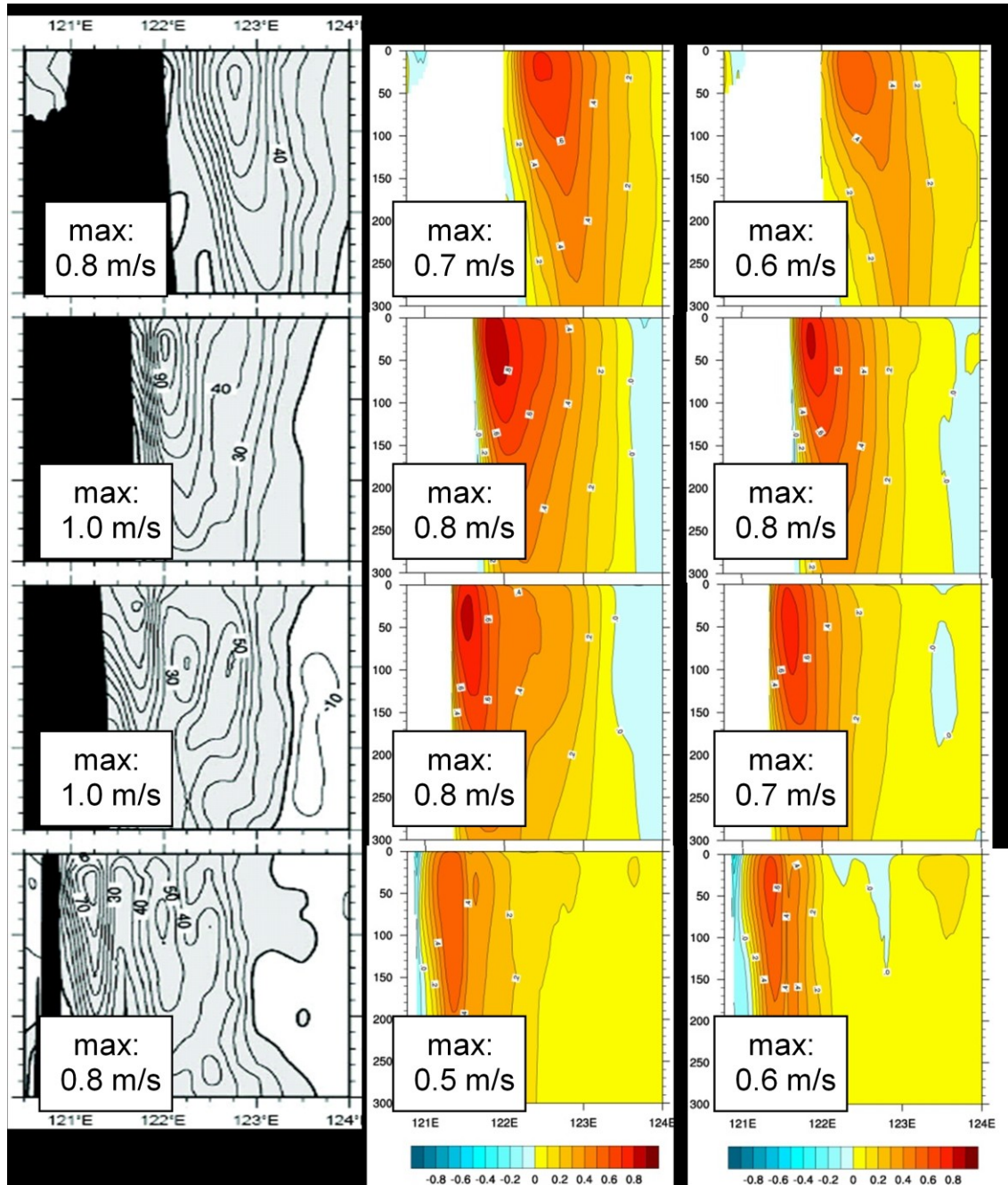


Figure 11: Mean meridional velocity (m/s) from the surface to 300 m from the Taiwan coast to 124°E at 25°N (top row), 24°N (second row), 23°N (third row) and 22°N (bottom row) from observations (left column) and the Relo NCOM system configured for the area in and around Luzon Strait forced with BCs from V3.0 (middle column) and V2.6 (right column). The Relo NCOM mean is computed over the time frame 3 September 2007 – 31 March 2008 and the observations are a 10-year composite of shipboard ADCP data from Liang et al. (2003). The contour interval is .1 m/s.

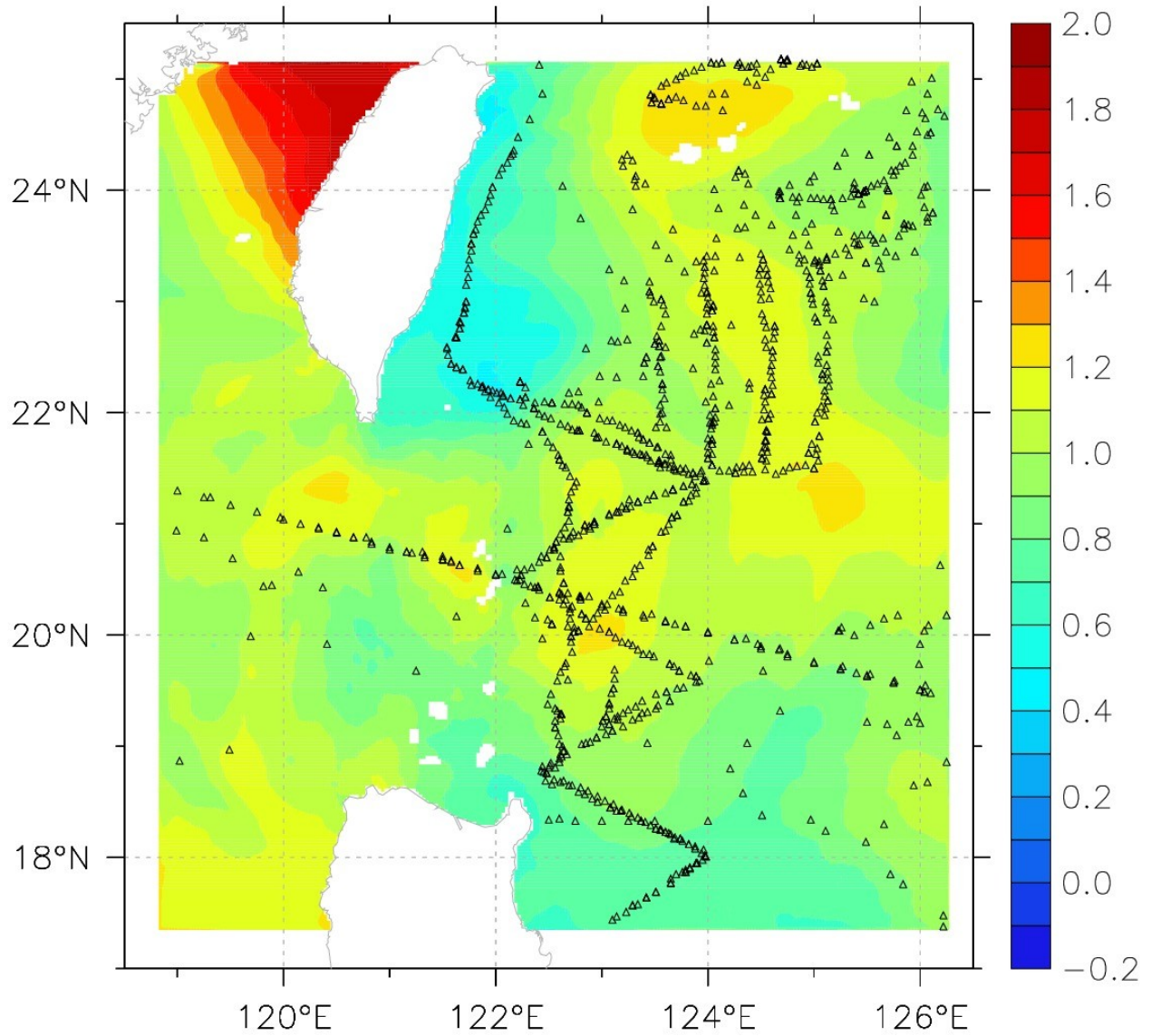


Figure 12: Positions of the 1144 unassimilated profiles (triangles) used in the error analyses of the Relo NCOM system configured for the area in and around Luzon Strait. The positions are overlain atop SSH (in meters) on 25 Sep 2007 for the simulation forced with V3.0 BCs.

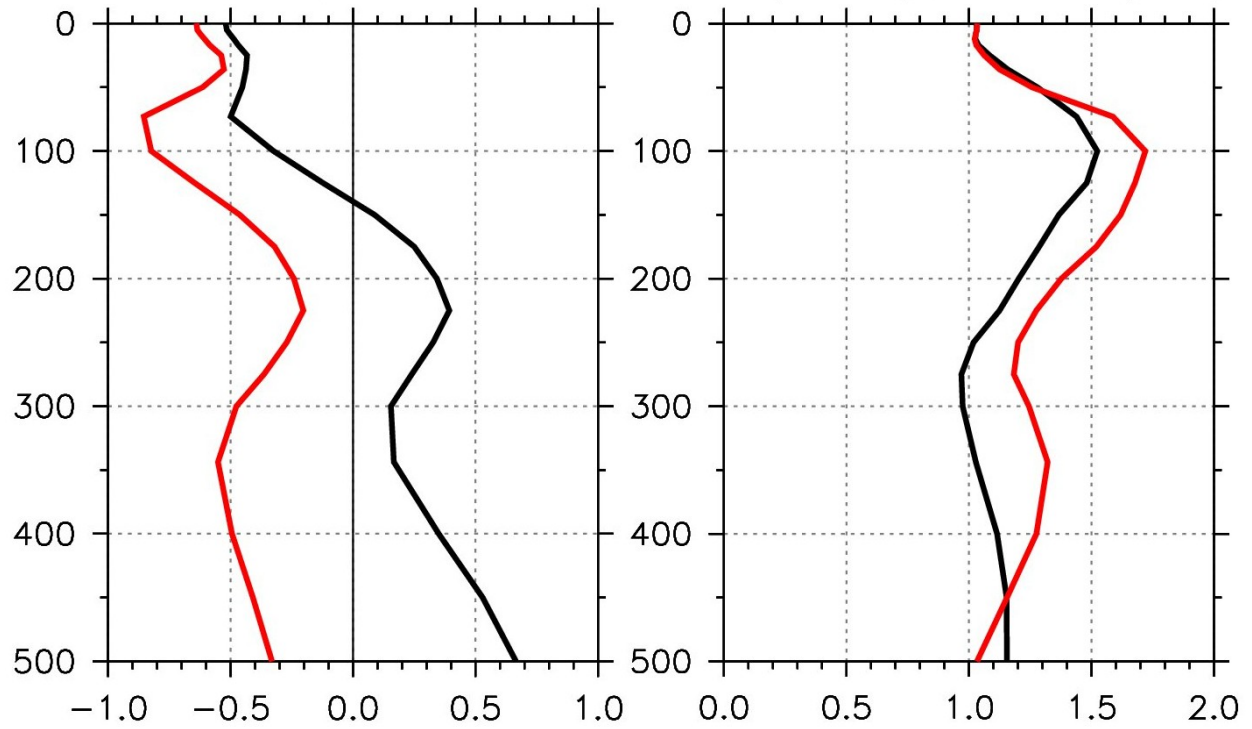


Figure 13: Temperature (°C) vs. depth error analysis in the upper 500 m against unassimilated profiles for Luzon Strait Relo NCOM forced with V3.0 BCs (black curves) and V2.6 BCs (red curves). The left panel is mean error and the right panel is RMSE. The number of near-surface unassimilated observations used is 1144 but that value decreases with depth since not all profiles extend down to 500 m depth.

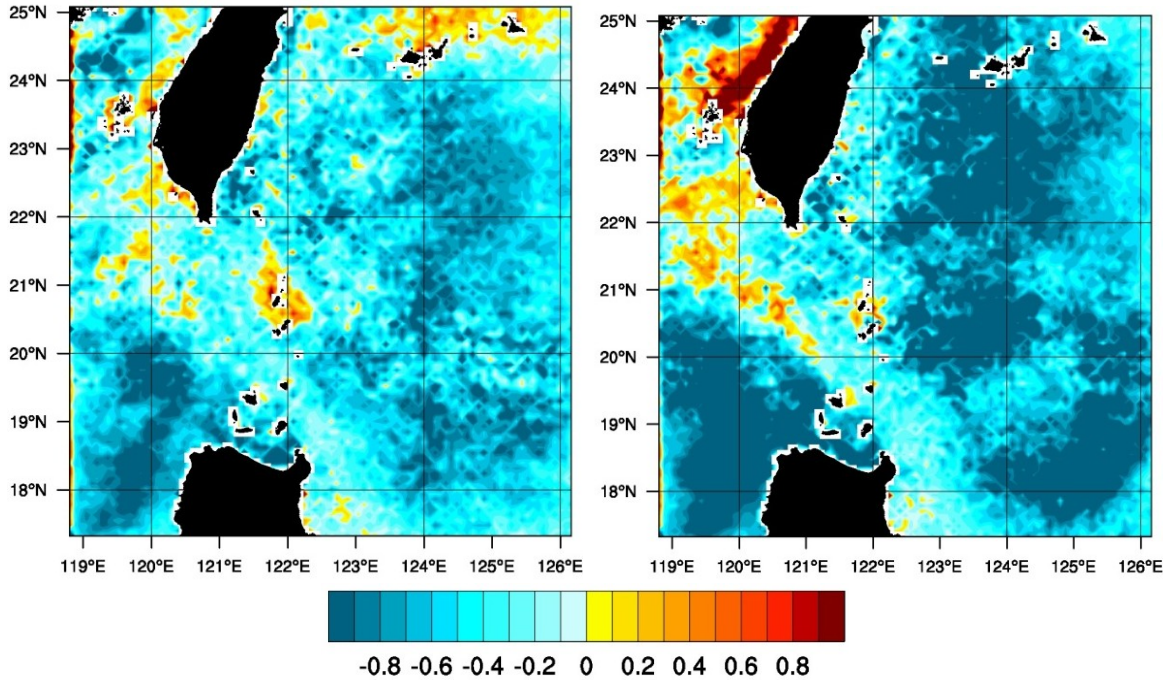


Figure 14: SST mean error (°C) for Luzon Strait Relo NCOM using V3.0 BCs (left) and V2.6 BCs (right) validated against ~173K MCSST observations at the nowcast time. The error is computed over the time frame 3 September 2007 – 31 March 2008. The average over the region for the simulation using V3.0 BCs is -0.52°C while the average for the simulation using V2.6 BCs is -0.67°C .

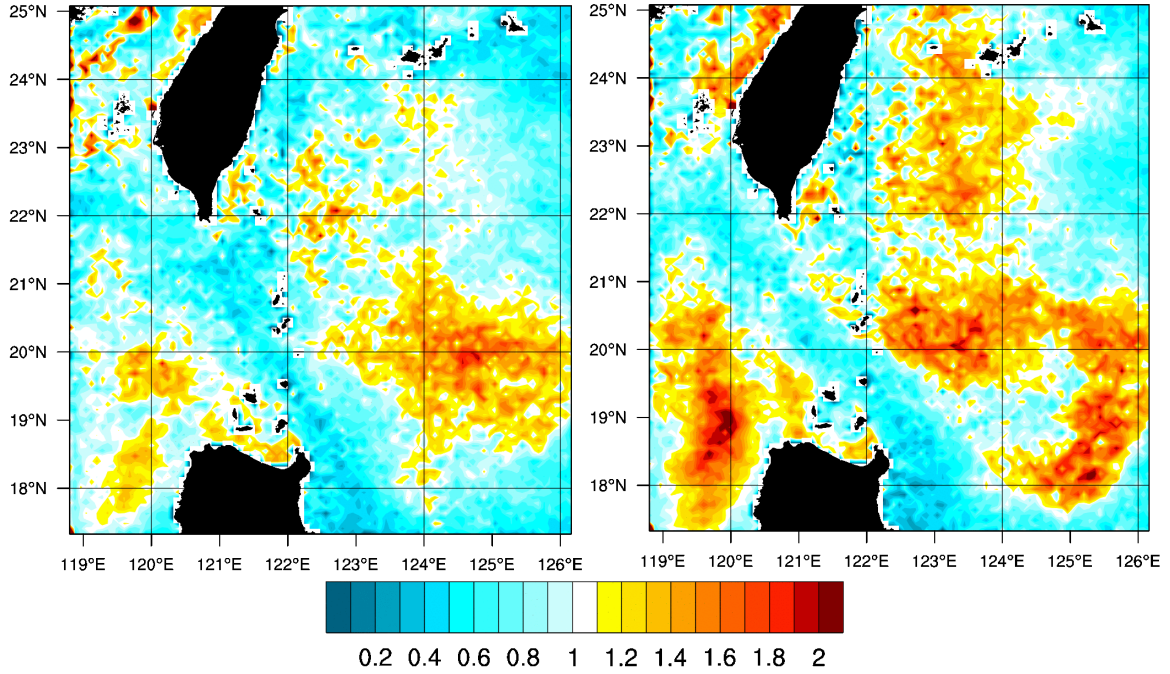


Figure 15: SST RMSE (°C) for Luzon Strait Relo NCOM using V3.0 BCs (left column) and V2.6 BCs (right column) validated against ~173K MCSST observations at the nowcast time. The error is computed over the time frame 3 September 2007 – 31 March 2008. The average over the region for the simulation using V3.0 BCs is 0.96°C while the average for the simulation using V2.6 BCs is 1.11°C.

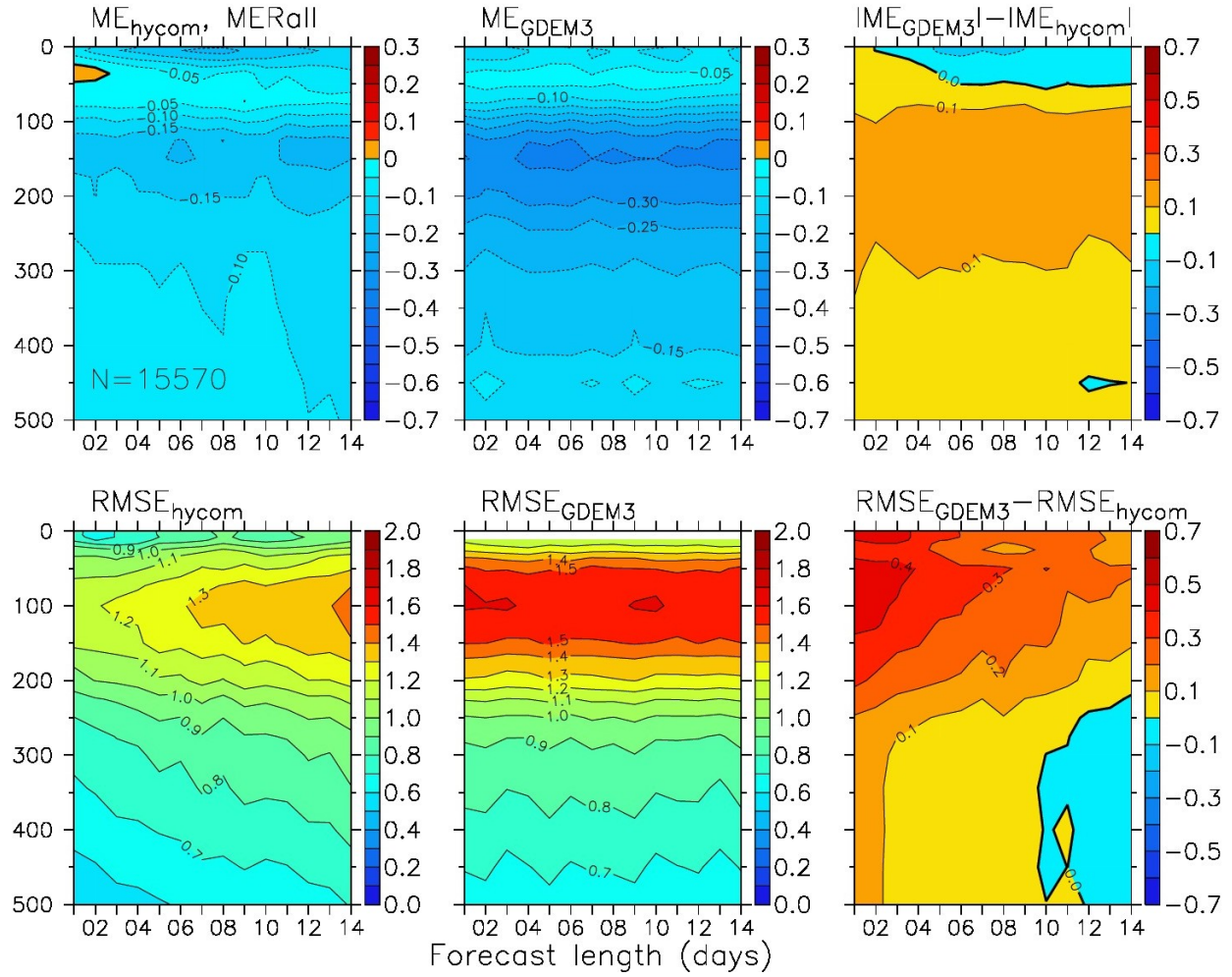


Figure 16: Temperature (°C) vs. depth error analysis in the upper 500 m as a function of forecast length based on 48 14-day forecasts of V3.0 for region MERall. The top row shows ME and the bottom row shows RMSE. The left column is for V3.0, the middle column is for the GDEM3 climatology and the right column is a difference of GDEM3 minus V3.0. In the right column, absolute value is used for mean error difference and the yellow-orange-red colors indicate that V3.0 has lower error whereas blue colors indicate GDEM3 has lower error. The average number of profiles used on each forecast day is 15570.

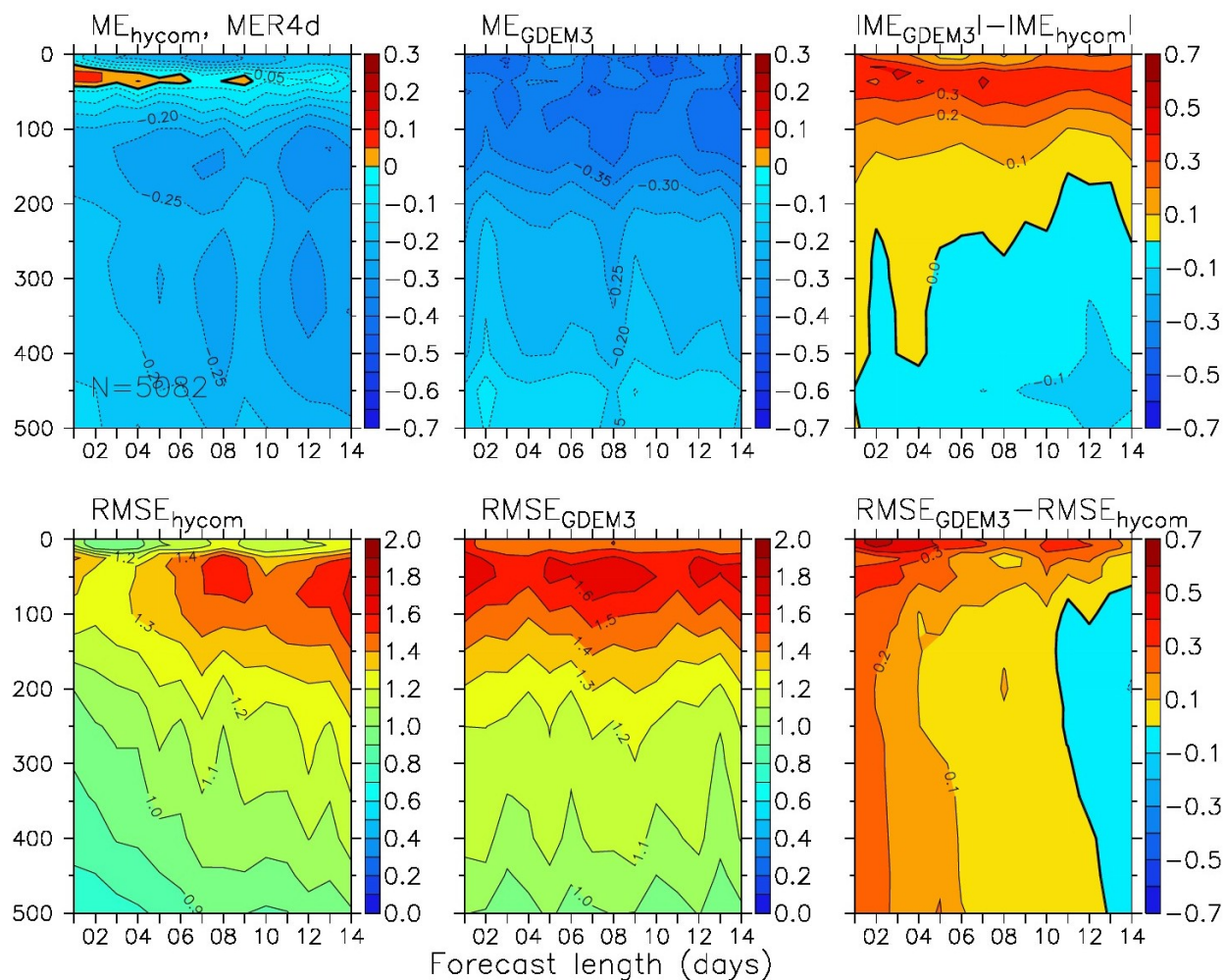


Figure 17: As in Figure 16 except for region MER4d. The average number of profiles used on each forecast day is 5082.

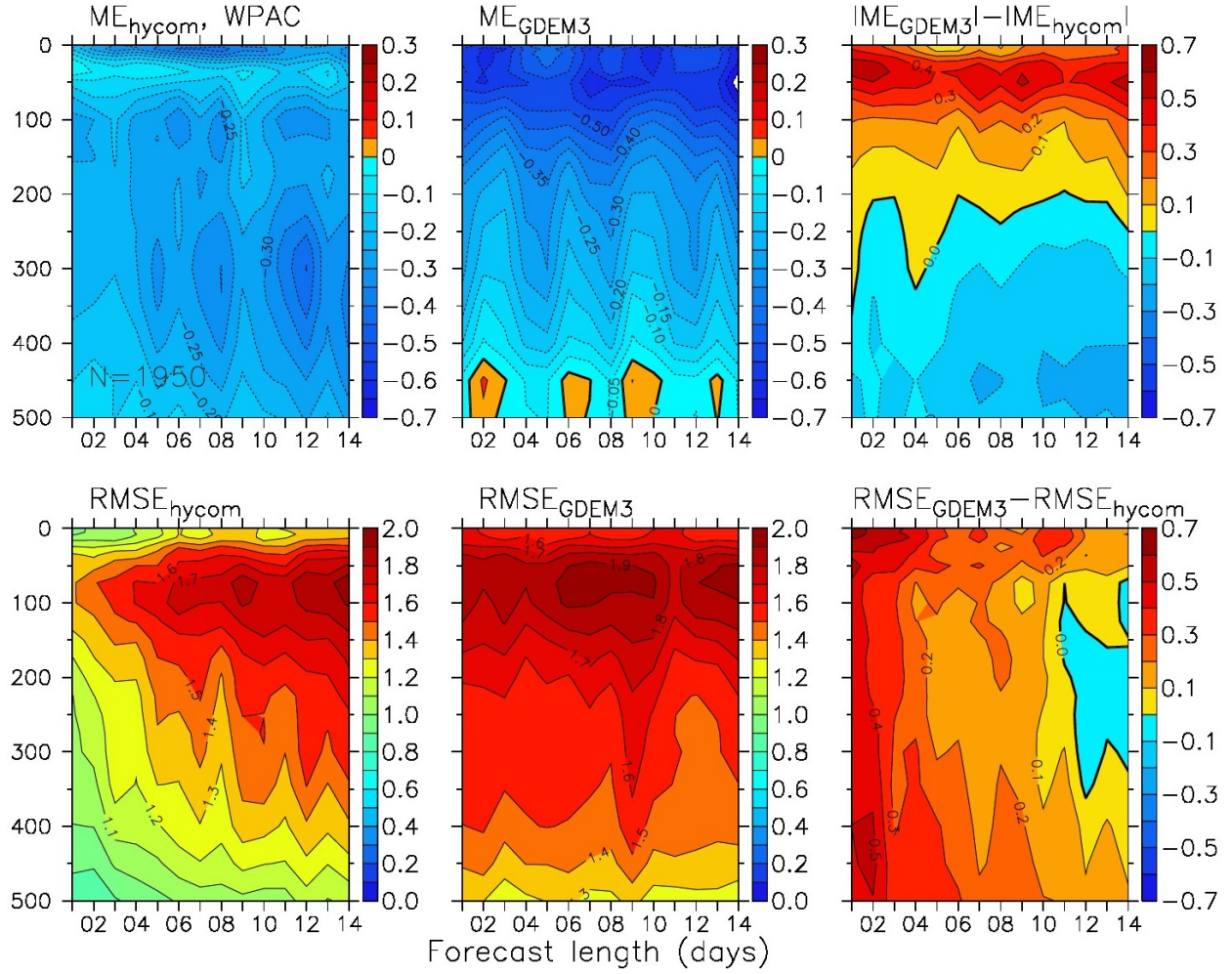


Figure 18: As in Figure 16 except for the western Pacific Ocean region. The average number of profiles used on each forecast day is 1950.

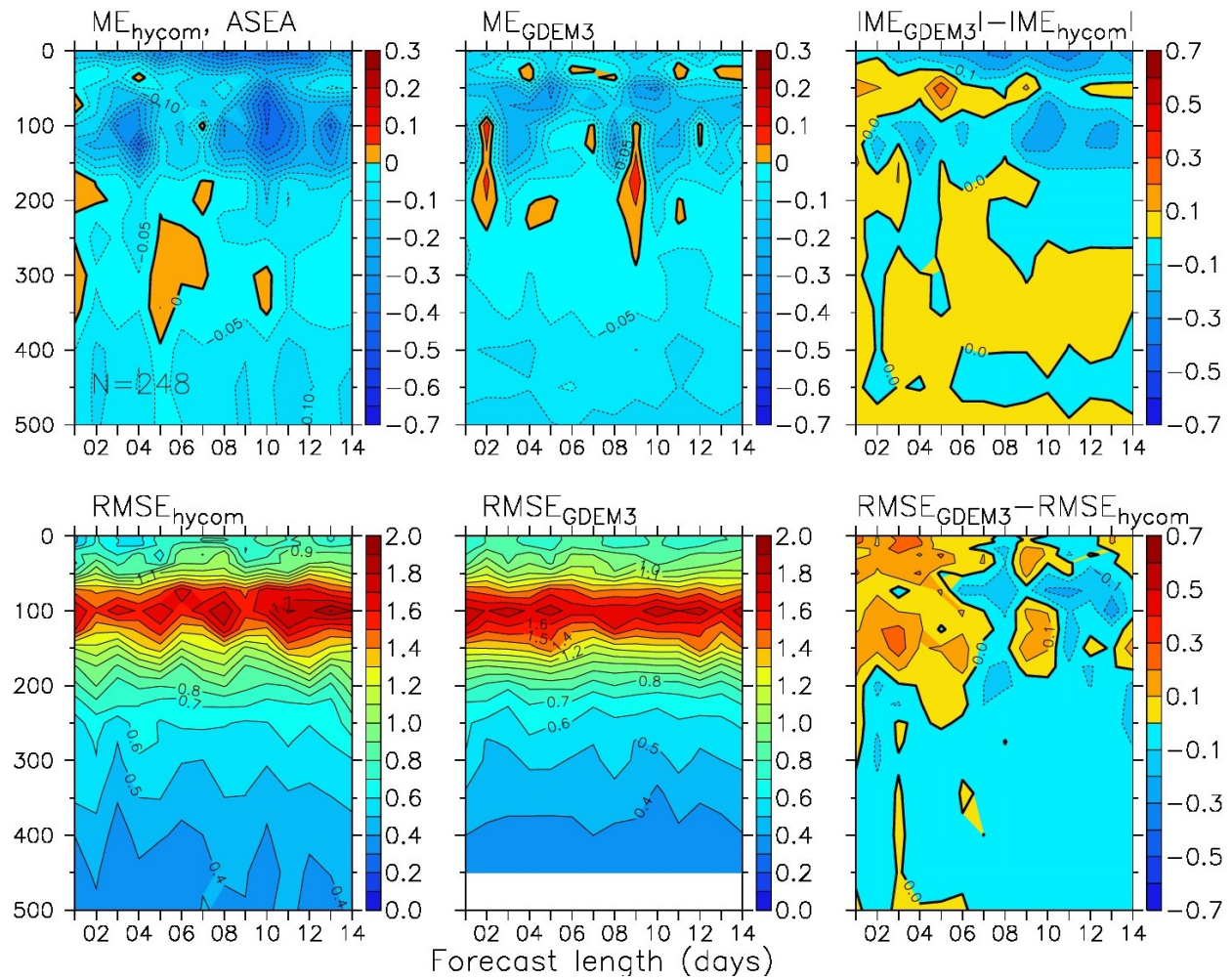


Figure 19: As in Figure 16 except for the Arabian Sea region. The average number of profiles used on each forecast day is 248.

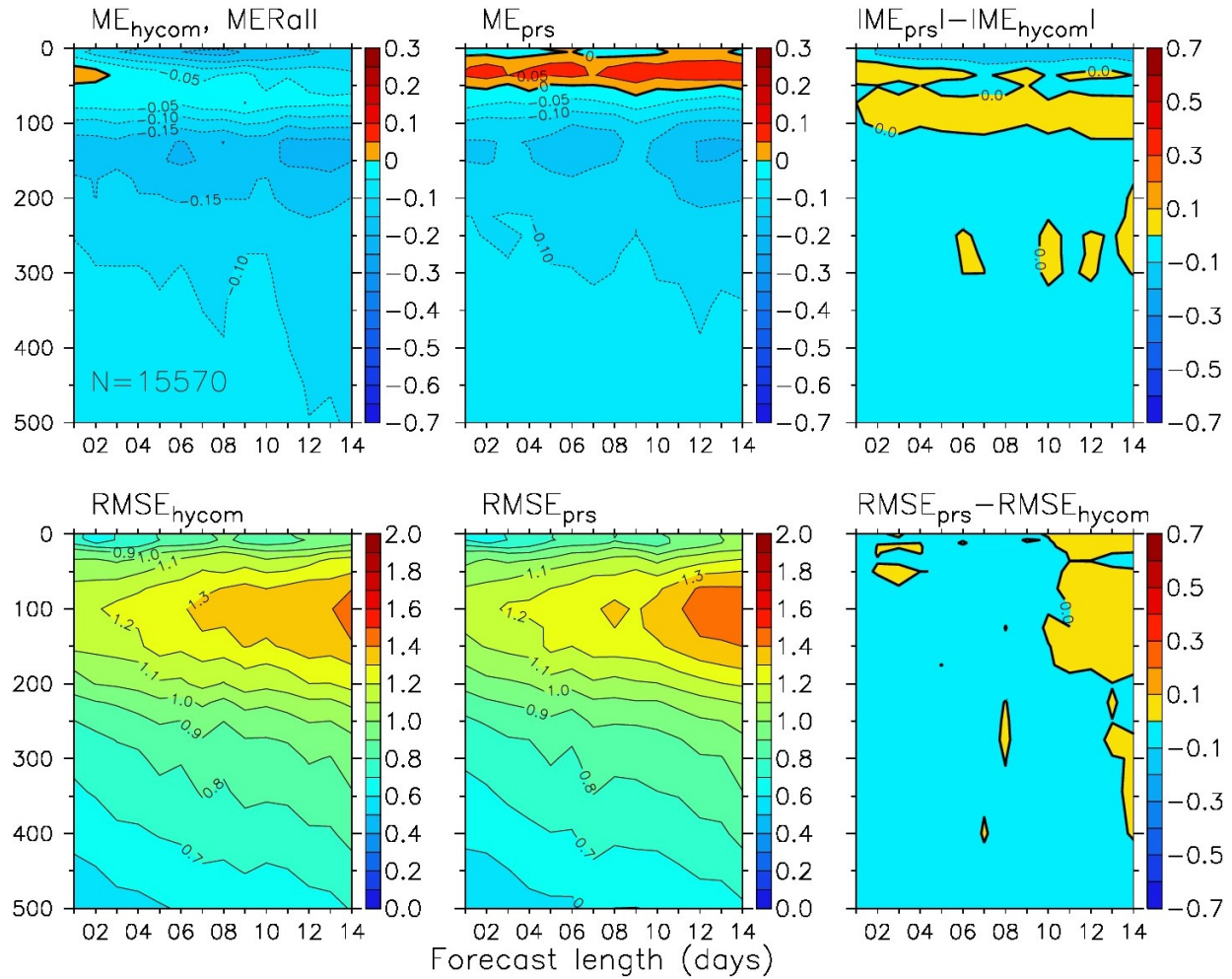


Figure 20: Temperature (°C) vs. depth error analysis in the upper 500 m as a function of forecast length based on 48 14-day forecasts of V3.0 for region MERall. The top row shows ME and the bottom row shows RMSE. The left column is for V3.0, the middle column is for persistence of the nowcast ocean state and the right column is a difference of persistence minus V3.0. In the right column, absolute value is used for mean error difference and the yellow-orange-red colors indicate that V3.0 has lower error whereas blue colors indicate persistence has lower error. The average number of profiles used on each forecast day is 15570.

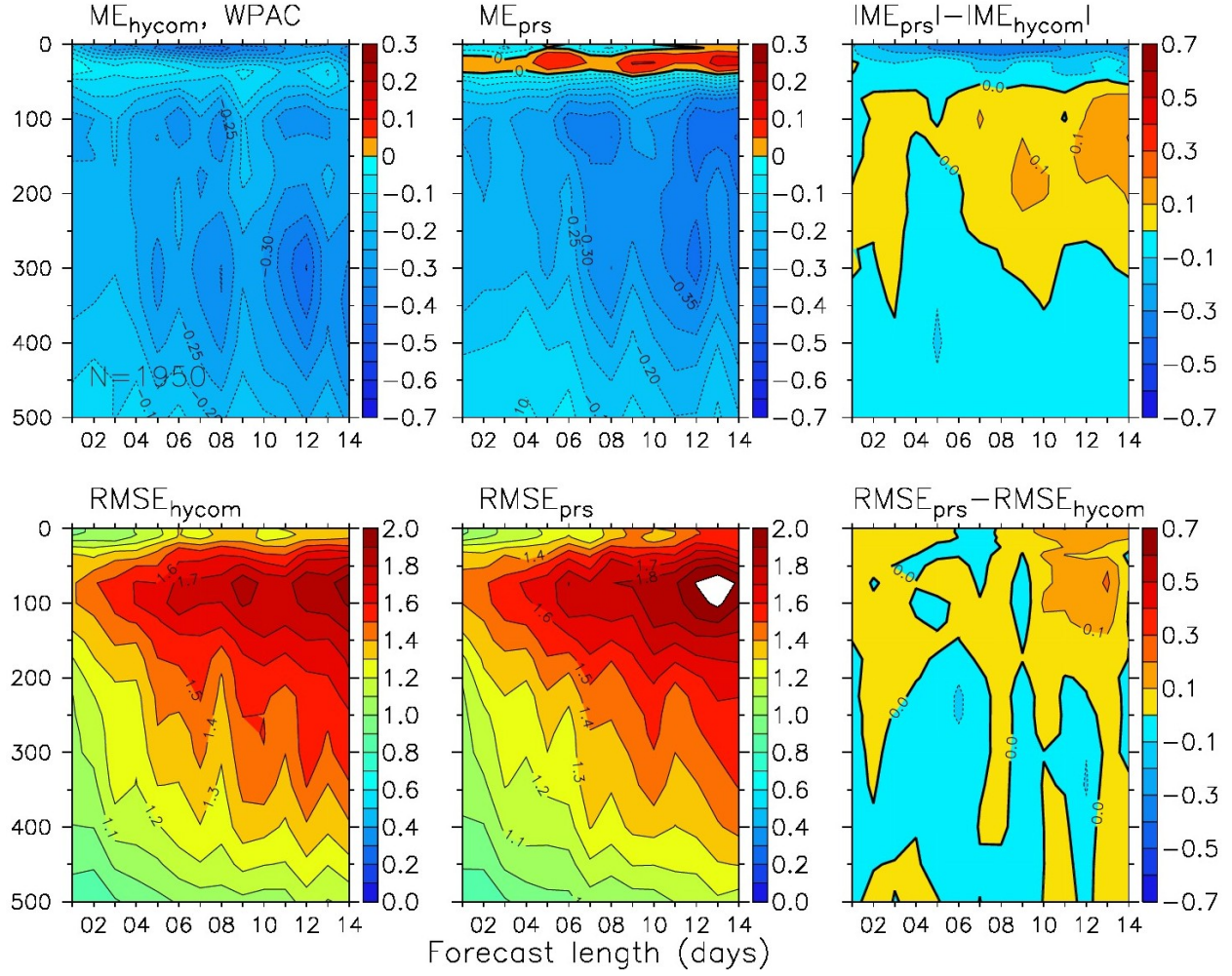


Figure 22: As in Figure 20 except for the western Pacific Ocean region. The average number of profiles used on each forecast day is 1950.

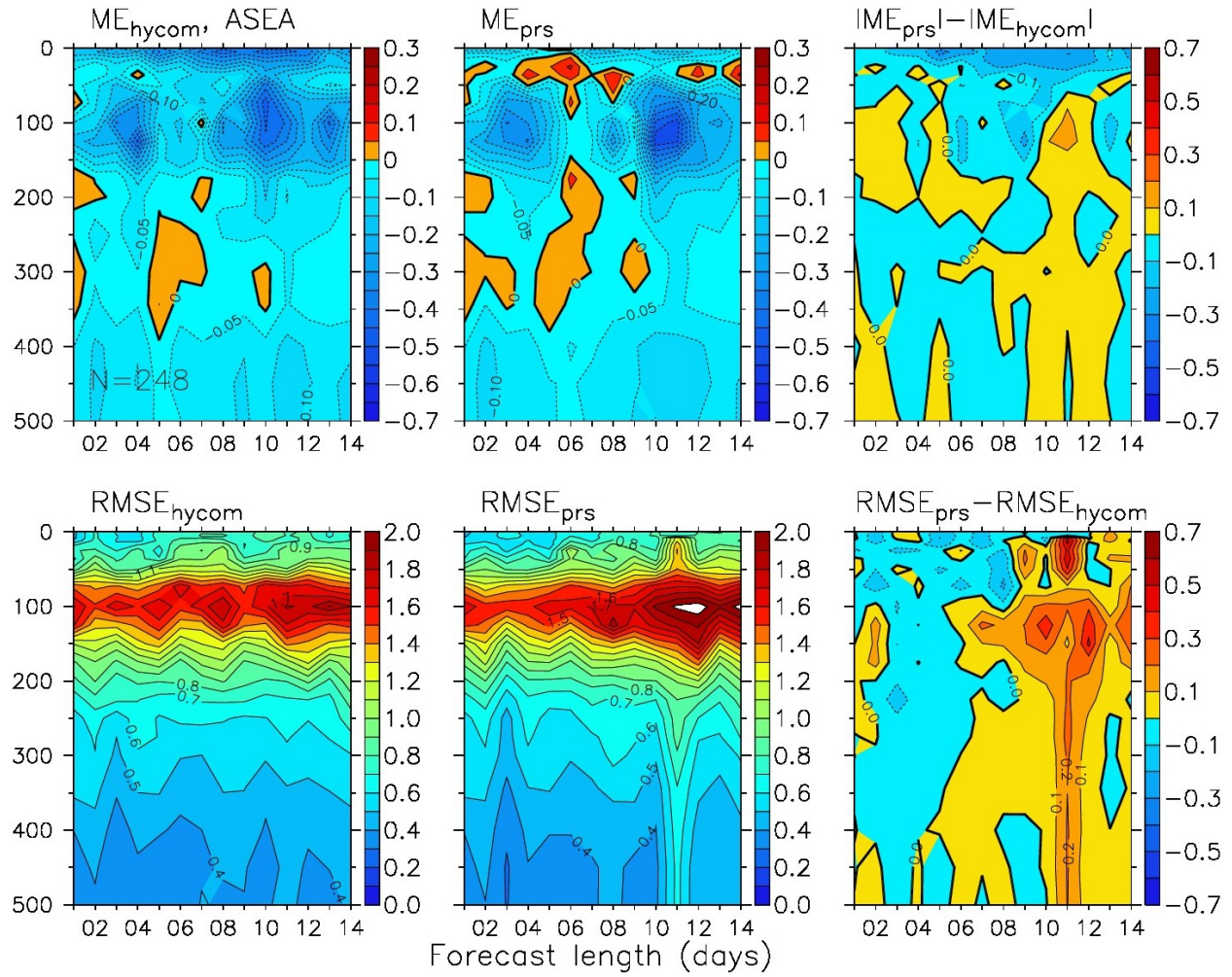


Figure 23: As in Figure 20 except for the Arabian Sea region. The average number of profiles used on each forecast day is 248.

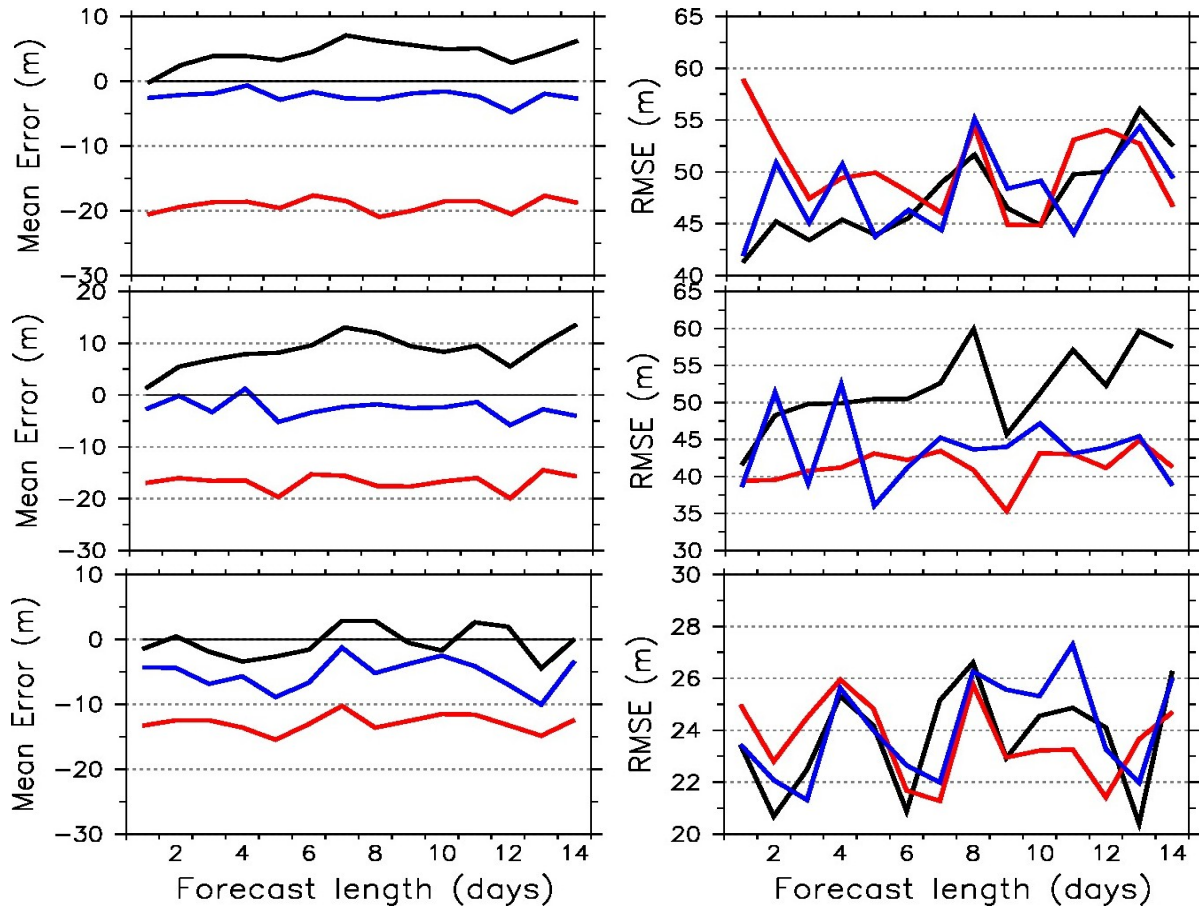


Figure 24: Error analysis of mixed layer depth (m) as a function of forecast length based on 48 14-day forecasts by V3.0 for regions MER4d (top), the western Pacific (middle) and the Arabian Sea (bottom). The left column shows mean error and the right column shows RMSE. The black curves are for V3.0 forecasts, the blue curves are for persistence of the nowcast ocean state and the red curves are for the GDEM3 climatology. Note the y-axis differs between most plots.

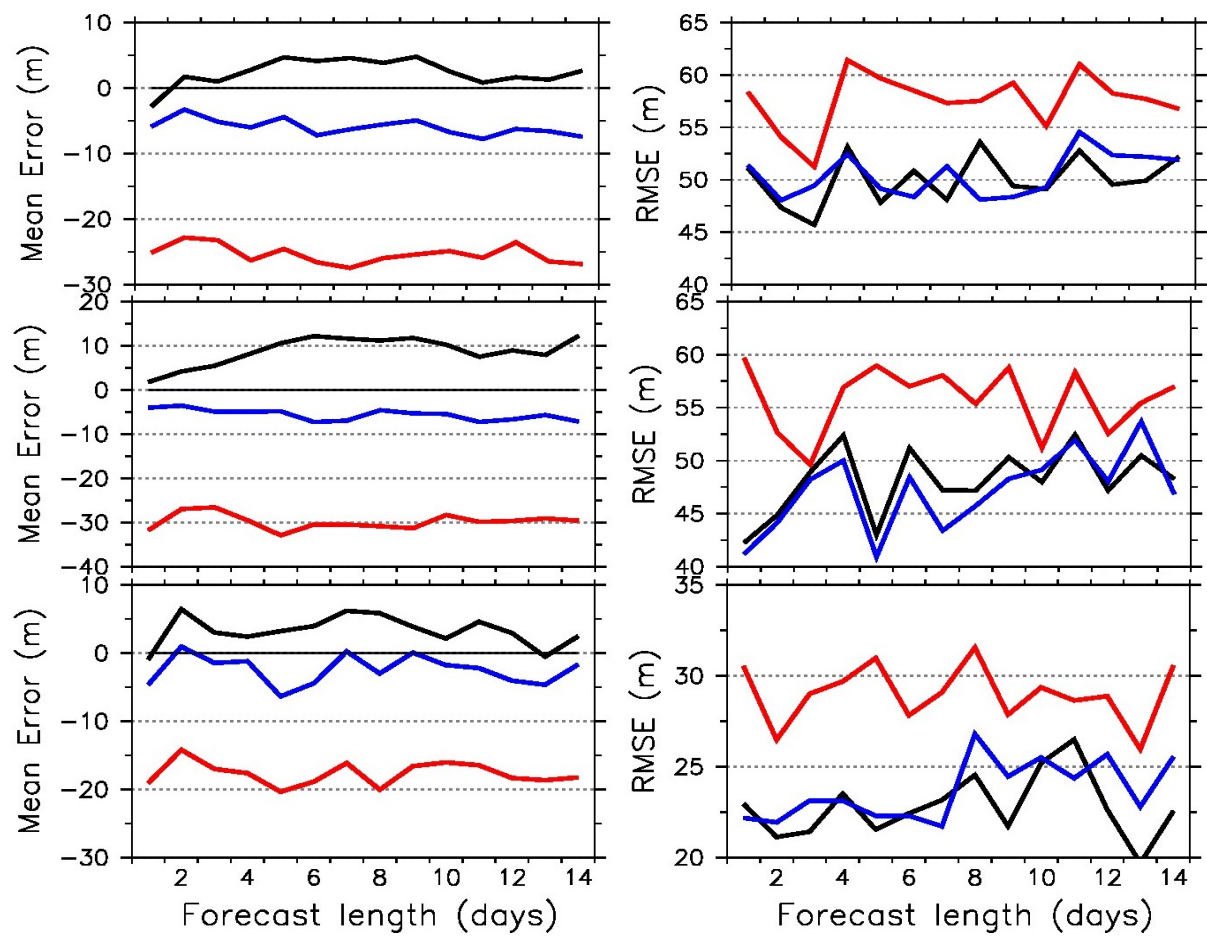


Figure 25: As in Figure 24 except for sonic layer depth.

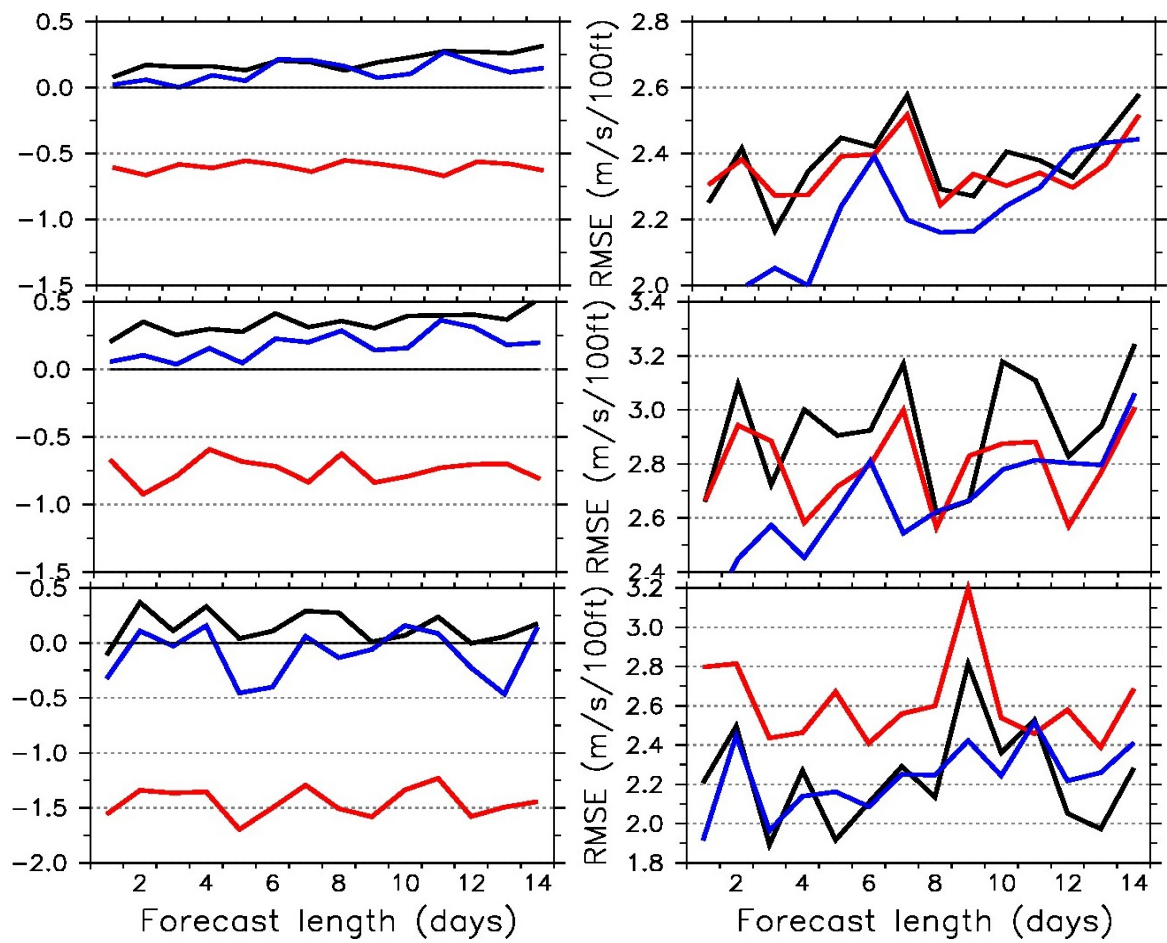


Figure 26: As in Figure 24 except for below layer gradient.

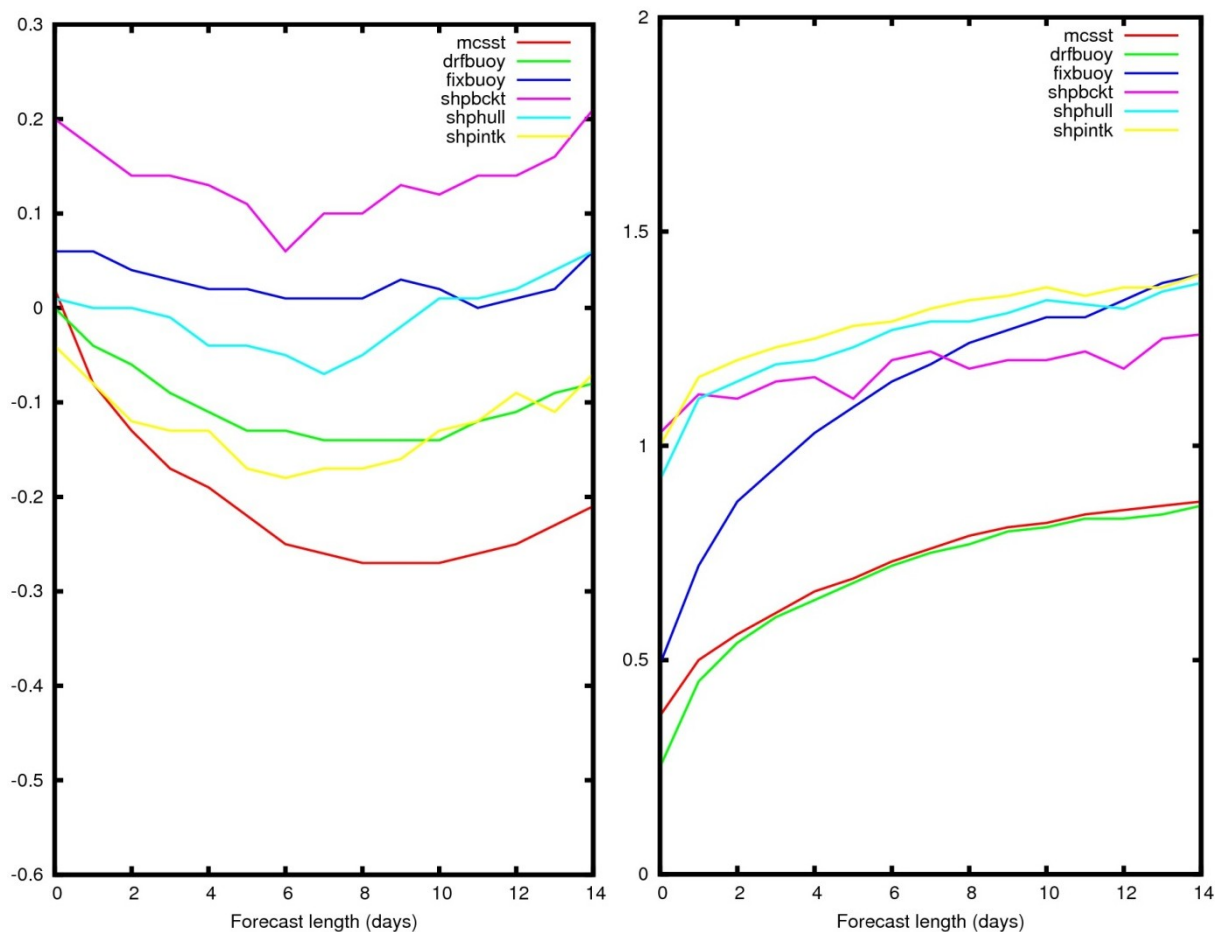


Figure 27: SST (°C) error analysis as a function of forecast length based on 48 14-day forecasts by V3.0 for the region spanning $\pm 45^\circ$ latitude. The left panel shows mean error and the right panel shows RMSE. The analysis is performed against different observational data types: satellite MCSST (red), drifting buoys (green), moored buoys (blue), ship bucket observations (magenta), ship hull observations (cyan) and ship intake observations (yellow). The approximate number of observations for each data type and each forecast are: 32 million – MCSST, 860K – drifting buoys, 172K – moored buoys, 2.4K – ship bucket, 14K – ship hull and 36K – ship intake.

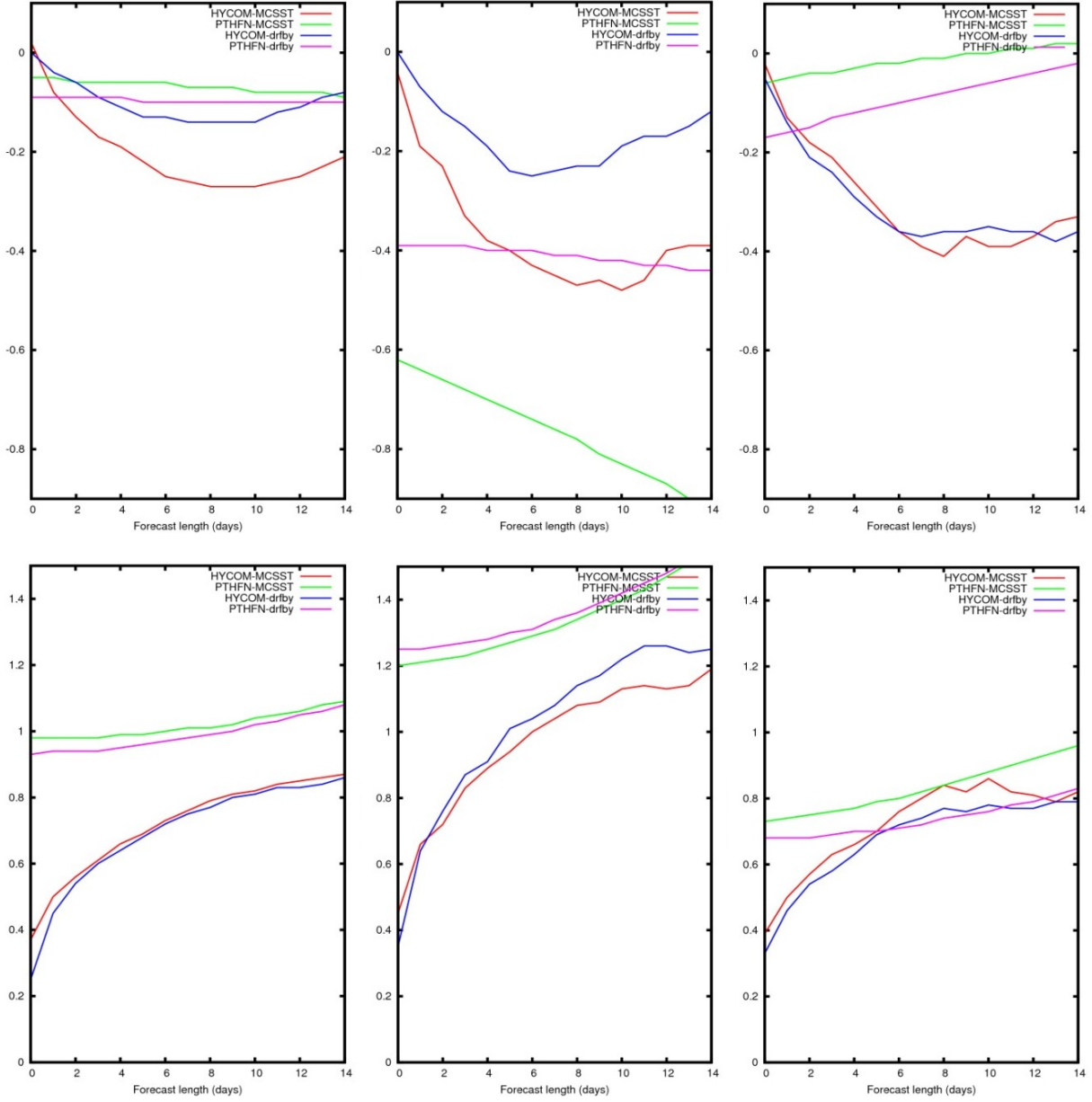


Figure 28: SST (°C) error analysis as a function of forecast length based on 48 14-day forecasts by V3.0 and for the 4 km Pathfinder SST climatology for the region spanning $\pm 45^\circ$ latitude (left column), the Northwest Pacific region (middle column) and the Arabian Sea region (right column). The top row shows mean error and the bottom row shows RMSE. The analysis is performed against two observational data types: V3.0 vs. satellite MCSST (red), V3.0 vs. drifting buoys (blue), Pathfinder vs. MCSST (green) and Pathfinder vs. drifting buoys (magenta).

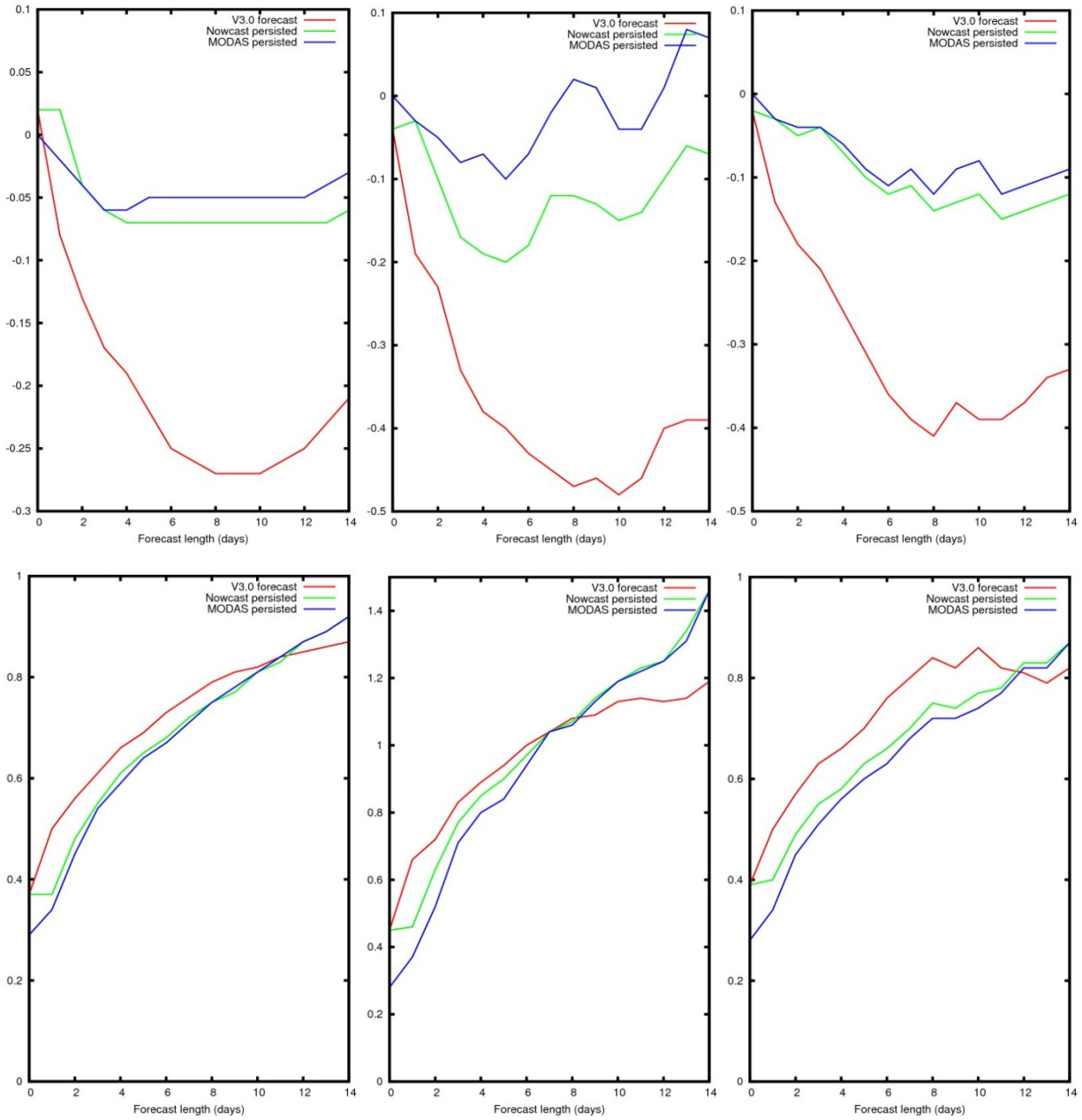


Figure 29: SST (°C) error analysis as a function of forecast length based on 48 14-day forecasts by V3.0 (red curves), for the V3.0 nowcast state persisted over the forecast period (green) and for the MODAS 2D SST analyses persisted over the forecast period (blue); all are against satellite MCSST observations. The top row shows mean error and the bottom row shows RMSE. The left column is for the region spanning $\pm 45^\circ$ latitude, the middle column for the Northwest Pacific region and the right column for the Arabian Sea region.

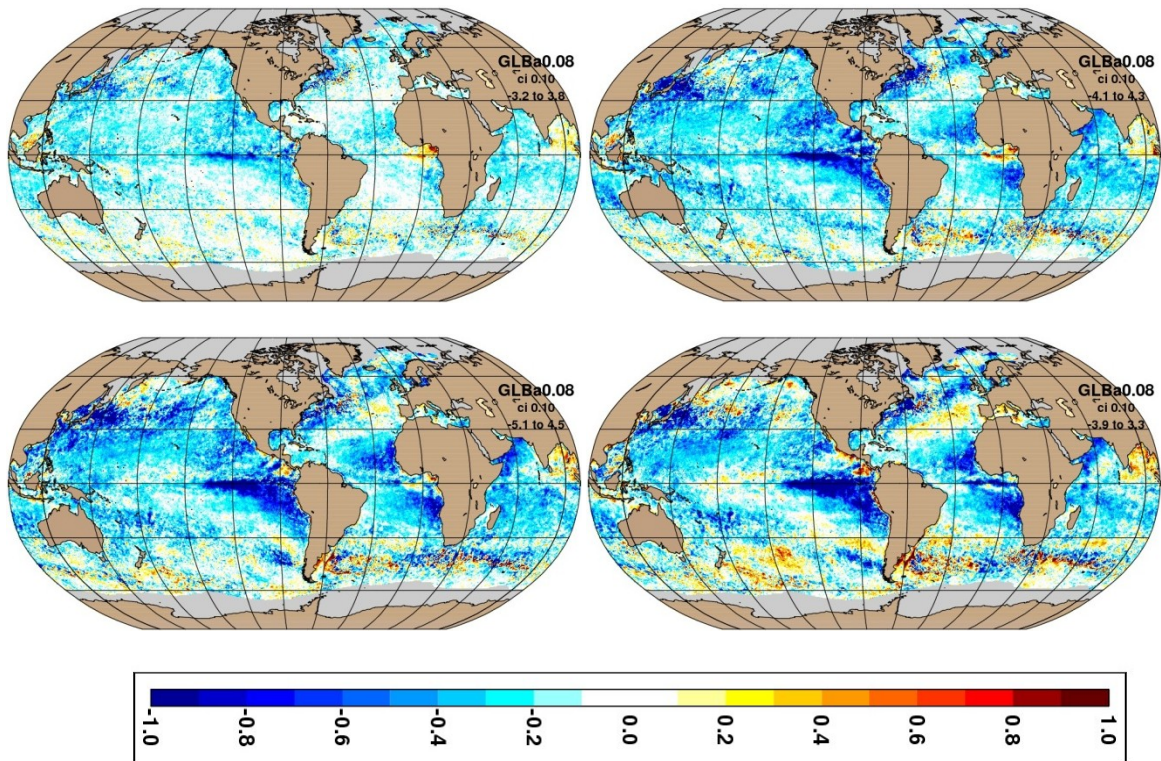


Figure 30: Spatial distribution of the V3.0 SST mean error (°C) compared against satellite MCSST observations for forecasts of 2 days (upper left), 6 days (upper right), 10 days (lower left) and 14 days (lower right). The gray areas near the poles are a 1982-2007 annual mean sea ice coverage mask from the Climate Diagnostics Center optimum interpolation SST analysis.

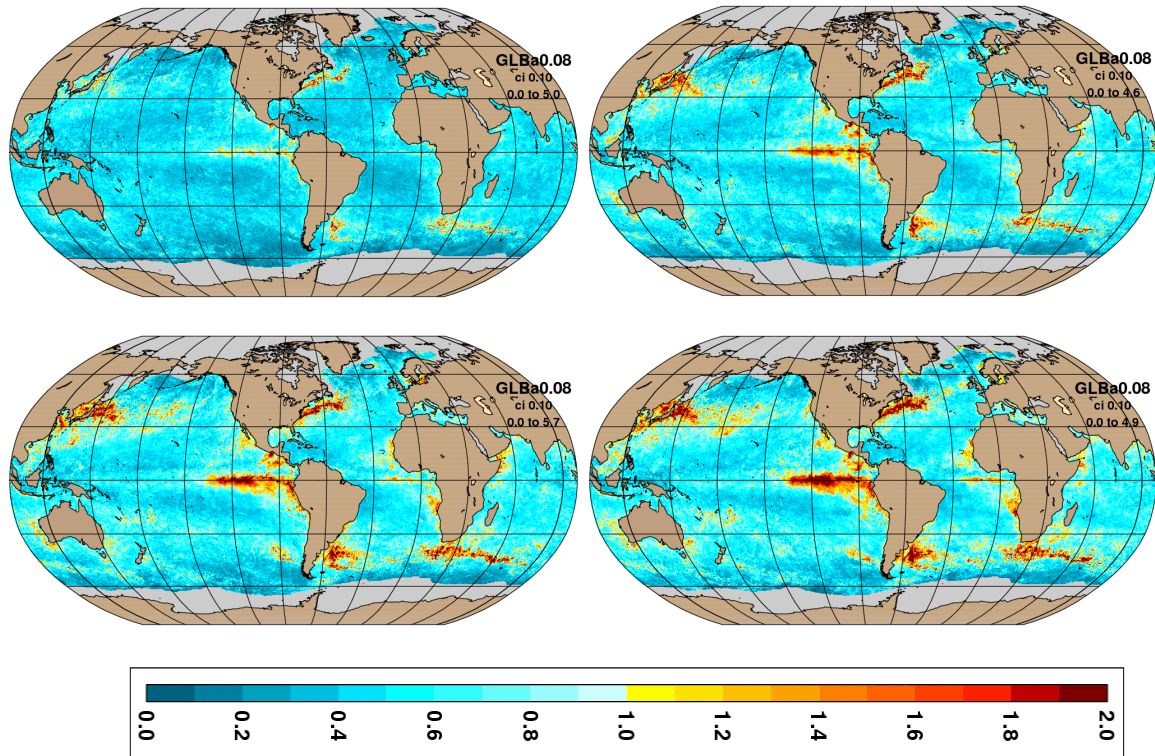


Figure 31: Spatial distribution of the V3.0 SST RMSE (°C) compared against satellite MCSST observations for forecasts of 2 days (upper left), 6 days (upper right), 10 days (lower left) and 14 days (lower right). The gray areas near the poles are a 1982-2007 annual mean sea ice coverage mask from the Climate Diagnostics Center optimum interpolation SST analysis.

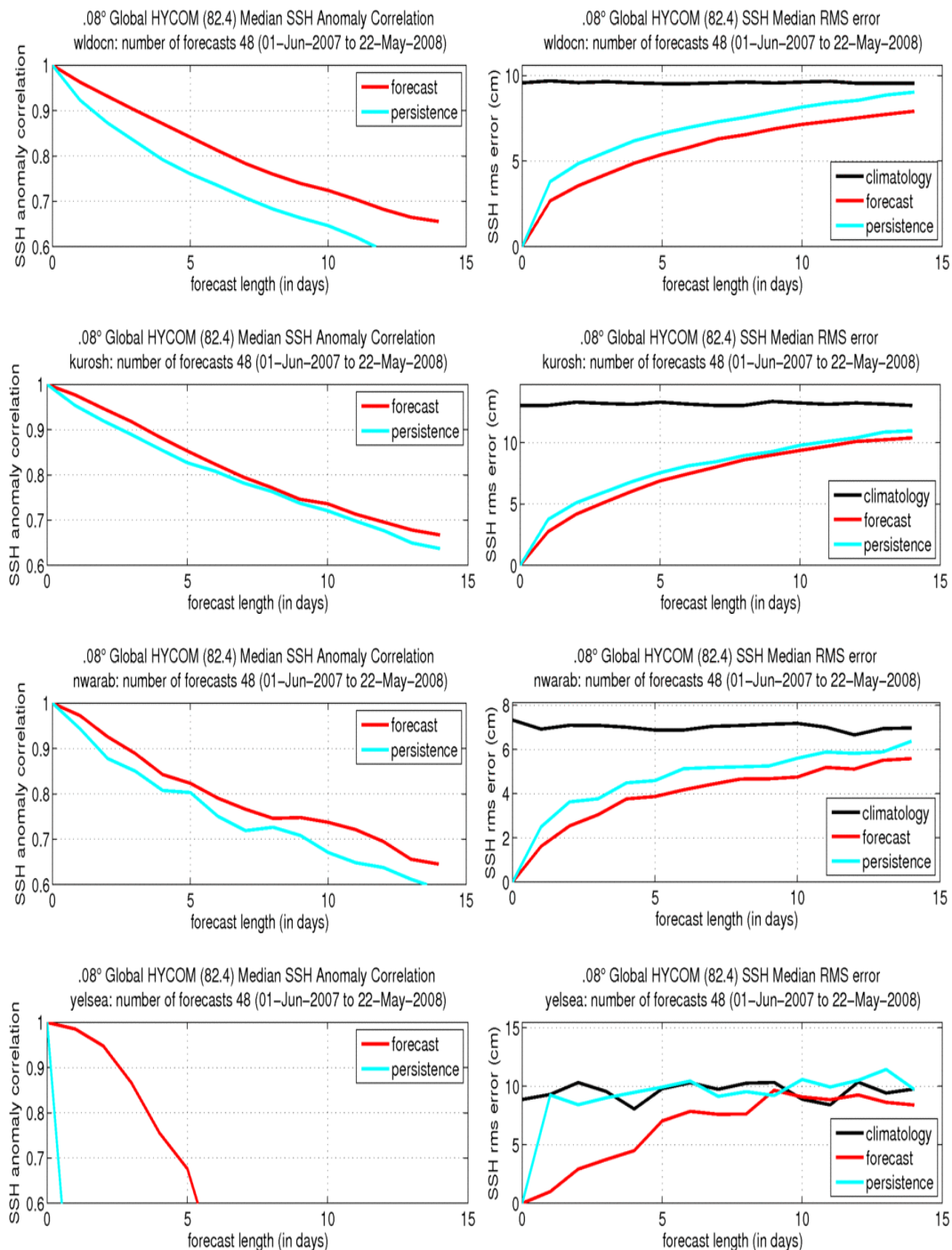


Figure 32: V3.0 median SSH anomaly correlation (left column) and median SSH RMSE (right column) against the verifying analysis as a function of forecast length for the global ocean (entire domain – top row), the Kuroshio (120-179°E, 21-55°N – second row), the northwest Arabian Sea (51-65°E, 15-26°N – third row) and the Yellow Sea (118-127°E, 30-42°N – bottom row). The red curves are V3.0 forecasts, the cyan curves are for persistence of the nowcast and the black curves of RMSE are for the hindcast annual mean.

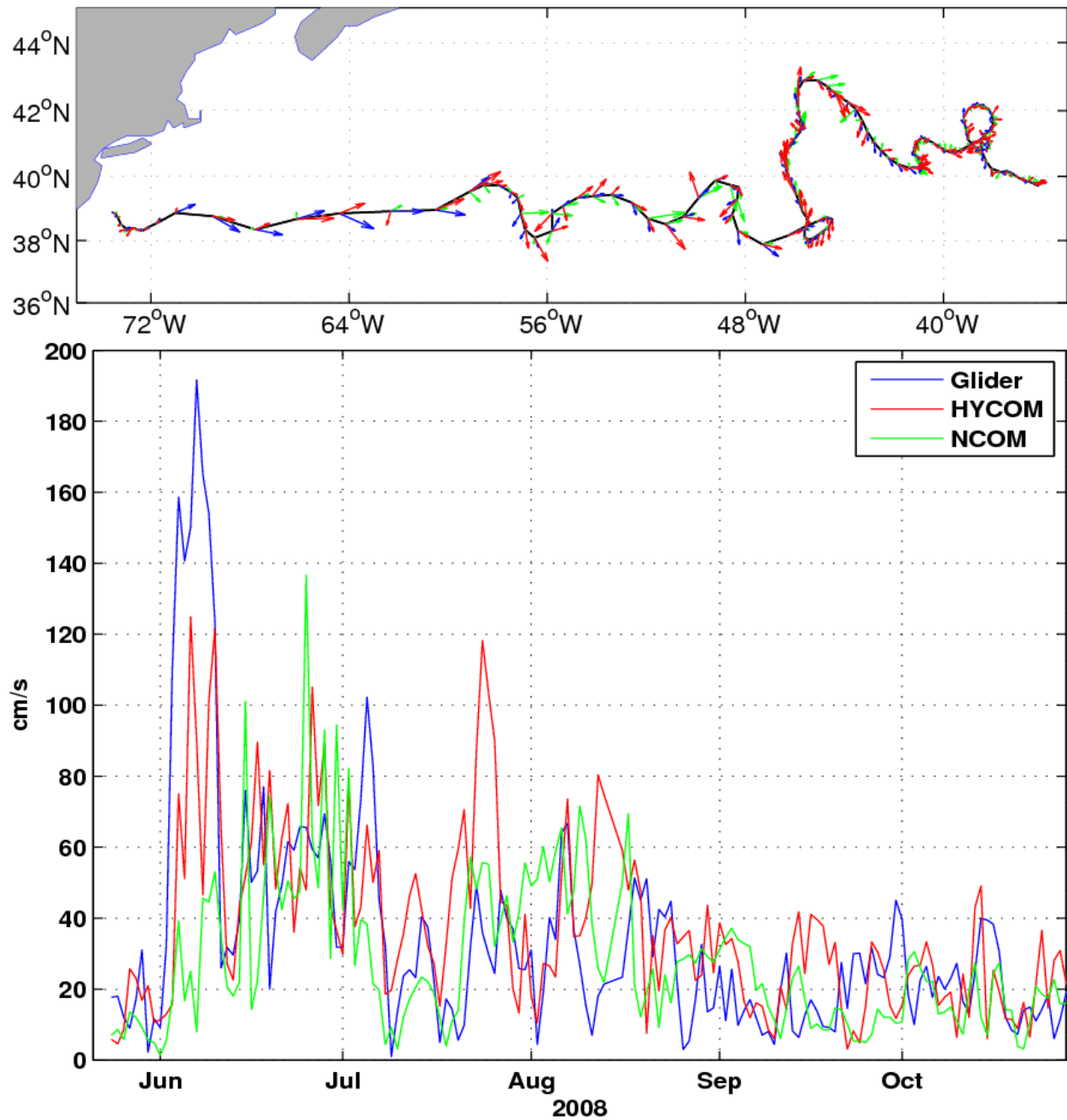


Figure 33: RU17 glider path (black line) and 100 m averaged current vectors (RU17 – blue arrows, V3.0 – red arrows and V2.6 – green arrows) sampled once per day nearest 00Z (top panel). One hundred meter depth averaged speed (cm/s) as a function of time (bottom panel) using the same color scheme.

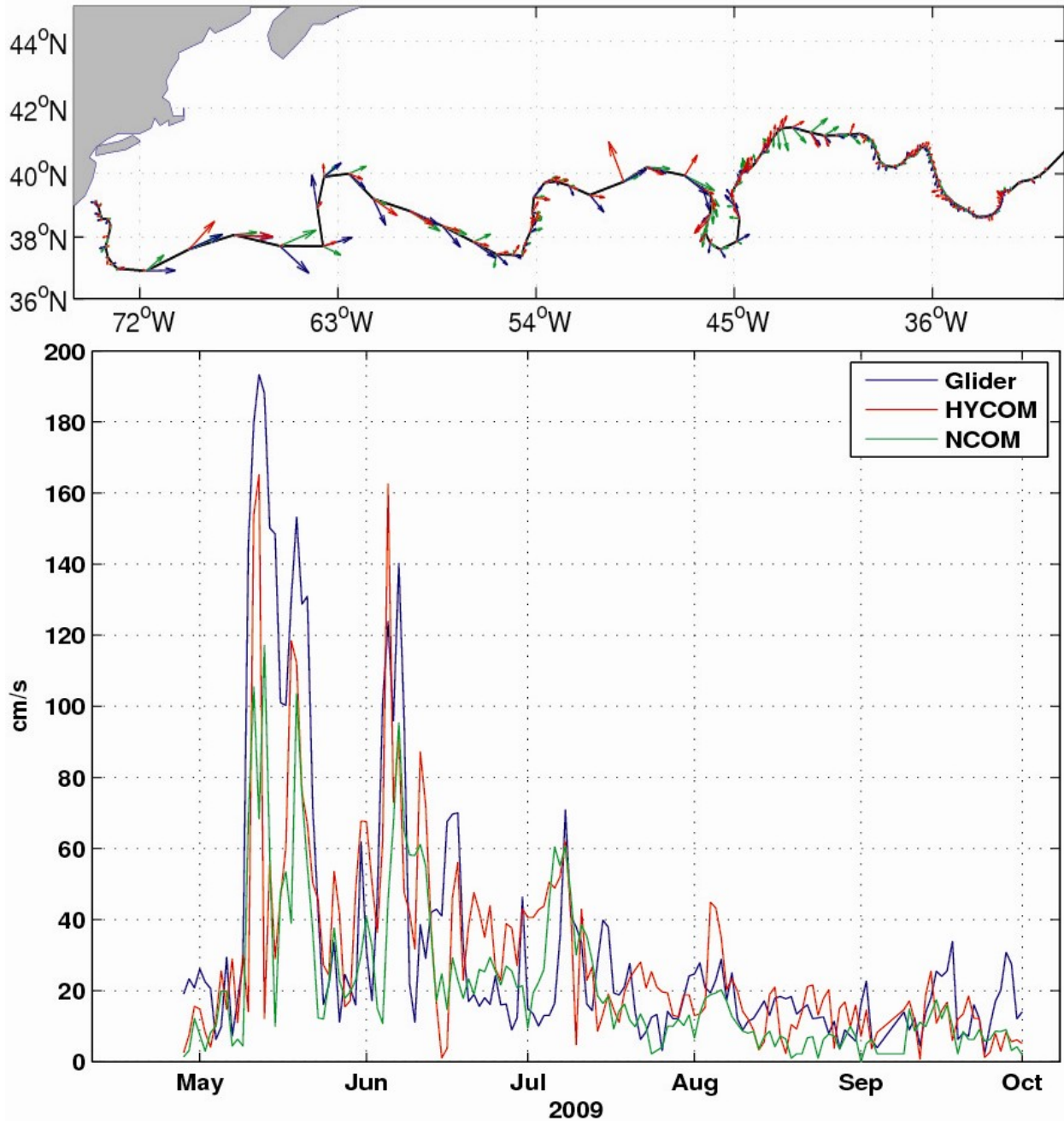


Figure 34: RU27 glider path (black line) and 150 m averaged current vectors (RU27 – blue arrows, V3.0 – red arrows and V2.6 – green arrows) sampled once per day nearest 00Z (top panel). One hundred fifty meter depth averaged speed (cm/s) as a function of time (bottom panel) using the same color scheme.

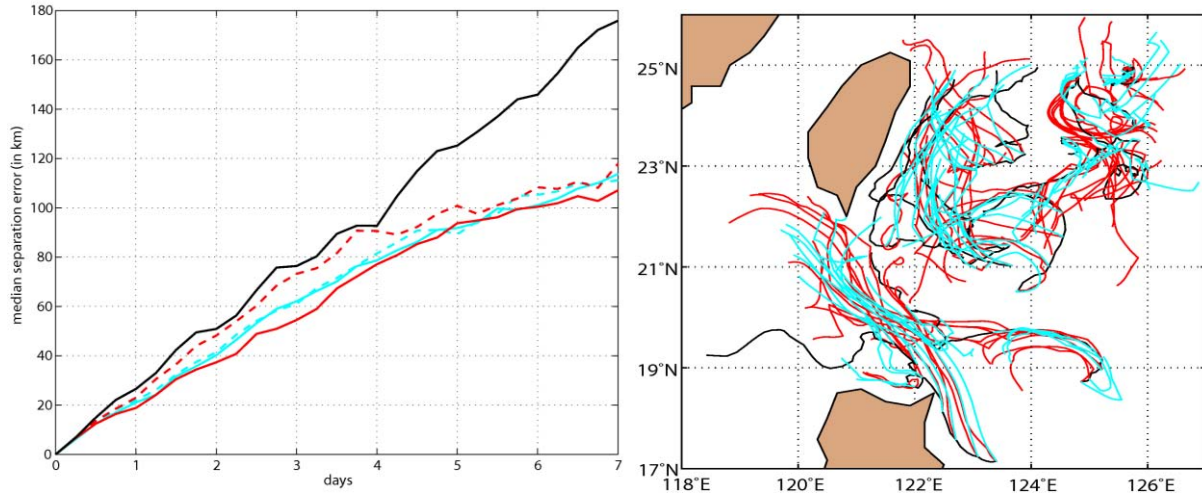


Figure 35: Median separation distance (km) as a function of time (left) between observed NOAA GDP drifting buoys and GOFS V3.0 (red), GOFS V2.6 (cyan) and no drift (black). The prediction system trajectories are computed using 6-hourly velocity fields (solid) and daily velocity fields (dashed). The analysis is performed over the period September 2007 – January 2008 using 97 trajectory pairs. The right panel shows the drifting buoy tracks used in the error analysis. The black lines are the observations, the red lines are GOFS V3.0 and the cyan lines are GOFS V2.6.

